



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 105953

TO: Karen A Lacourciere
Location: CM-1/11D09/11E12
Art Unit: 1635
Wednesday, October 15, 2003

Case Serial Number: 09/915543

From: Edward Hart
Location: Biotech-Chem Library
CM1-6B02
Phone: 305-9203

edward.hart@uspto.gov

Search Notes

Examiner Lacourciere,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC-Biotech/ChemLib

105953

From: Lacourciere, Karen
Sent: Tuesday, October 14, 2003 3:13 PM
To: STIC-Biotech/ChemLib
Subject: Sequence search request 09/915,543

Please perform the following sequence searches for 09/915,543 in the commercial databases, pre-grant pubs and pending files (interference)

- 1) Residues 177-204 of SEQ ID NO:15
- 2) Residues 349-383 of SEQ ID NO:15
- 3) Residues 199-392 of SEQ ID NO:15

Thank-you!

Karen A. Lacourciere Ph.D.

CM1 11D09 GAU 1635
(703) 308-7523
mailbox 11E12

RECEIVED
OCT 14 2003
STIC

Edward Han
Technical Info. Specialist
STIC/Biotech
CM1 0832 Tel 308-9209

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 10/15/03
Date Completed: 10/15/03
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 3
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: DT
WWW/Internet: _____
Other (specify): _____



STIC SEARCH RESULTS

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or contact:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art found, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not** found:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 – Circ. Desk



Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 17.214 Seconds
(without alignments)
258.182 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204
Perfect score: 136
Sequence: 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136	100.0	1426	23 AAB71229	Human legless homo
2	136	100.0	1435	22 ABB11808	Human BCL9 homology
3	129	94.9	140	22 AAO55855	Human polypeptide
4	98	72.1	1429	22 ABB58779	Drosophila melanog
5	98	72.1	1464	23 AAB71228	D. melanogaster lg
6	94	69.1	1494	23 AAU78460	Mouse beta-catenin
7	91	66.9	114	23 ABP06595	Human ORFX protein
8	67	49.3	320	23 AAU78461	Mouse beta-catenin
9	65	47.8	113	22 AAO07544	Human polypeptide

10	52	38.2	360	21 AAG33446	Zea mays protein f
11	52	38.2	448	21 AAG33445	Zea mays protein f
12	52	38.2	509	21 AAG33444	Zea mays protein f
13	51	37.5	1049	22 ABB60387	Drosophila melanog
14	50	36.8	130	23 ABB89793	Human polypeptide
15	50	36.8	365	22 ABG15088	Novel human diagno
16	50	36.8	621	23 AAE15740	Human aminoacyl-tr
17	50	36.8	631	21 AAB43285	Human ORFX ORF3049
18	50	36.8	1063	23 ABB08919	Human aminoacyl tr
19	49	36.0	187	23 ABP66271	Bifidobacterium lo
20	49	36.0	984	24 ABJ25889	Aspergillus fumiga
21	49	36.0	1058	24 ABJ26489	Aspergillus fumiga
22	48	35.3	498	21 AAG17973	Arabidopsis thalia
23	48	35.3	498	21 AAG49364	Arabidopsis thalia
24	48	35.3	609	21 AAG17972	Arabidopsis thalia
25	48	35.3	609	21 AAG49363	Arabidopsis thalia
26	48	35.3	630	21 AAG17971	Arabidopsis thalia
27	48	35.3	630	21 AAG49362	Arabidopsis thalia
28	47	34.6	886	21 AAG40431	Arabidopsis thalia
29	47	34.6	905	21 AAG39800	Arabidopsis thalia
30	47	34.6	1095	21 AAG40430	Arabidopsis thalia
31	47	34.6	1114	21 AAG39799	Arabidopsis thalia
32	47	34.6	1168	21 AAG40429	Arabidopsis thalia
33	47	34.6	1187	21 AAG39798	Arabidopsis thalia
34	47	34.6	2639	20 AAW73476	Grapevine leafroll
35	45.5	33.5	341	22 ABB70889	Drosophila melanog
36	45	33.1	401	22 ABG04882	Novel human diagno
37	44.5	32.7	119	21 AAG08467	Arabidopsis thalia
38	44.5	32.7	137	21 AAG51008	Arabidopsis thalia
39	44.5	32.7	152	21 AAG08466	Arabidopsis thalia
40	44.5	32.7	160	21 AAG08465	Arabidopsis thalia
41	44.5	32.7	170	21 AAG51007	Arabidopsis thalia
42	44.5	32.7	170	21 AAY77970	A. thaliana enviro
43	44.5	32.7	177	21 AAG51006	Arabidopsis thalia
44	44.5	32.7	447	22 AAG82903	S. epidermidis ope
45	44.5	32.7	450	22 AAG83066	S. epidermidis ope

ALIGNMENTS

RESULT 1

AAB71229
ID AAB71229 standard; Protein; 1426 AA.

XX AAB71229;

XX 18-NOV-2002 (first entry)

XX Human legless homologue lgs/bcl9 protein.

XX Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW tissue proliferation; tumour; cytosolic; cellular disorder; colon;
KW blood disorder; cancer; breast; head and neck cancer; thyroid;
KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.

XX Homo sapiens.

XX US2002086986-A1.

XX 04-JUL-2002.

XX 27-JUL-2001; 2001US-0915543.

XX 28-JUL-2000; 2000US-221502P.

XX (BASL/) BASLER K.

XX (BRUN/) BRUNNER E.

XX (FROE/) FROESCH B.

XX (KRAM/) KRAMPS T.

XX (PETE/) PETER O.

PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;

XX WPI; 2002-635689/68.
DR N-PSDB; AAF88467.
XX
PT Novel polypeptide useful in therapeutic method for treating disorders
PT of cell fate such as cell differentiation or cell proliferation -
XX
XX Example II; Fig 8B; 4lpp; English.
XX
CC This invention describes a novel polypeptide sharing one or more
CC homologous amino acid domains with the legless (lgs) protein, a
CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC involved in the formation and maintenance of spatial arrangements
CC and proliferation of tissues during development, and in the formation
CC and growth of many human tumours. The products of the invention have
CC cytoskeletal activity and can be used to treat cellular disorders, blood
CC disorders and cancers caused by over-stimulation of the Wnt pathway,
CC where the cancerous condition is colon, breast, head and neck, brain,
CC thyroid, medulloblastoma or skin cancer. The product could also be used
CC to promote tissue regeneration and repair. This sequence represents the
CC human legless (lgs) protein homologue lgs/bcl9 described in the
CC disclosure of the invention.
XX
SQ Sequence 1426 AA;
Query Match 100.0%; Score 136; DB 23; Length 1426;
Best Local Similarity 100.0%; Pred. No. 2.9e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
Db 177 VYVFSTEMANKAAEAVLKQGVETIVSFH 204
RESULT 2
ABB11808
ID ABB11808 standard; peptide; 1435 AA.
XX
AC ABB11808;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human BCL9 homologue, SEQ ID NO:2178.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03800.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX

DR WPI; 2001-457740/49.
DR N-PSDB; ABA09052.
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
XX Claim 20; Page 256-257; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
XX
SQ Sequence 1435 AA;
Query Match 100.0%; Score 136; DB 22; Length 1435;
Best Local Similarity 100.0%; Pred. No. 2.9e-12;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
Db 217 VYVFSTEMANKAAEAVLKQGVETIVSFH 244
RESULT 3
AAO05855
ID AAO05855 standard; Protein; 140 AA.
XX
AC AAO05855;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 19747.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX

OS Homo sapiens.
XX
XX WO200164835-A2.
XX
XX PD 07-SEP-2001.
XX
XX PF 26-FEB-2001; 2001WO-US04927.
XX
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Drmanac RT;
XX
XX DR WPI; 2001-514838/56.
XX DR N-PSDB; AAI85786.
XX
XX PT Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders -
XX
XX PS Claim 20; SEQ ID NO 19747; 1399pp + Sequence Listing; English.
XX
XX CC The invention relates to human polynucleotides (AAI93841) and
XX CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 140 AA;

Query Match 94.9%; Score 129; DB 22; Length 140;
Best Local Similarity 92.9%; Pred. No. 2.3e-12;
Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
Db |||||:|||||:|||||:|||||:|||||
59 VYVFSTEMANKAAKAVLKQGVETIVSFH 86

RESULT 4
ABB58779
ID ABB58779 standard; Protein; 1429 AA.
XX
XX AC ABB58779;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster polypeptide SEQ ID NO 3129.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX
XX OS Drosophila melanogaster.
XX
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX

XX (PEKE) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL02882.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX
XX PS Disclosure; SEQ ID NO 3129; 2lpp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 1429 AA;

Query Match 72.1%; Score 98; DB 22; Length 1429;
Best Local Similarity 57.1%; Pred. No. 2.8e-06;
Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
Db ::|||::||| ||::|| ||::|||
323 IFVFSTQLANKGAESVLSGQFTIIAYH 350

RESULT 5
AAB71228
ID AAB71228 standard; Protein; 1464 AA.
XX
XX AC AAB71228;
XX
XX DT 18-NOV-2002 (first entry)
XX
XX DE D. melanogaster lgs protein.
XX
XX KW Legless; fruitfly; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
XX KW tissue proliferation; tumour; cytostatic; cellular disorder; colon;
XX KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
XX KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX
XX OS Drosophila melanogaster.
XX
XX PN US2002086986-A1.
XX
XX PD 04-JUL-2002.
XX
XX PF 27-JUL-2001; 2001US-0915543.
XX
XX PR 28-JUL-2000; 2000US-221502P.
XX
XX PA (BASL/) BASLER K.
XX PA (BRUN/) BRUNNER E.
XX PA (FROE/) FROESCH B.
XX PA (KRAM/) KRAMPS T.
XX PA (PETE/) PETER O.
XX
XX PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
XX DR WPI; 2002-635689/68.
XX DR N-PSDB; AAF88466.
XX

PT Novel polypeptide useful in therapeutic method for treating disorders
PT of cell fate such as cell differentiation or cell proliferation -
XX
PS Example II; Fig 2; 4lpp; English.

This invention describes a novel polypeptide sharing one or more homologous amino acid domains with the legless (lgs) protein, a downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway involved in the formation and maintenance of spatial arrangements and proliferation of tissues during development, and in the formation and growth of many human tumours. The products of the invention have cytostatic activity and can be used to treat cellular disorders, blood disorders and cancers caused by over-stimulation of the Wnt pathway, where the cancerous condition is colon, breast, head and neck, brain, thyroid, medulloblastoma or skin cancer. The product could also be used to promote tissue regeneration and repair. This sequence represents the *Drosophila melanogaster* (fruitfly) legless (lgs) protein described in the disclosure of the invention.

Sequence	1464	AA:
50		

Query Match	72.1%;	Score 98;	DB 23;	Length 1464;
Best Local Similarity	57.1%;	Pred. No. 2.9e-06;		
Matches 16; Conservative	9;	Mismatches 3;	Indels 0;	Gaps 0;

QY 1 VVVFSTEMANKAAEAVLKGQVETIVSEH 28
:::||||| |::| | |::|:
DB 318 IEFVSTOLANKAESVLGFOFOTIAYH 345

RESULT 6
AAU78460

AAU78460 standard: Protein; 1494 AA.

AA
AC
AAU78460:

02-JUL-2002 (first entry)

Mouse beta-catenin nuclear localised protein.

Mouse; beta-catenin nuclear localised protein; cancer;
KW
gene therapy; EST; expressed sequence tag.
KW

OS Mus musculus.

PN WO200224738-A1.

PD 28-MAR-2002.

19-SEP-2001; 2001WO-JP08140.

22-SEP-2000: 2000JP-0287876.

PA (KYOW) KYOWA HAKKO KOGYO KK.

PI Akivama T., Adachi S:

WPI: 2002-330014/36.

XX
XX N-PSDB; HBK47031.

aa	New beta-catenin nuclear localised protein for diagnosis and treatment
pt	of diseases associated with nuclear localisation of beta-catenin e.g.
PT	cancer -

xx PS claim 1: page 81-88: 113pp: Japanese.

The invention relates to a beta-catenin nuclear localised protein
 and DNA encoding the protein. The protein and encoding DNA are
 applicable in diagnosis and treatment of diseases associated with
 nuclear localisation of beta-catenin e.g. cancer, including gene
 therapy. The present sequence represents the amino acid sequence of
 mouse beta-catenin nuclear localised protein.

SQ Sequence 1494 AA;

Query Match 69.1%; Score 94; DB 23; Length 1494;
Best Local Similarity 60.7%; Pred. No. 1.3e-05;
Matches 17; Conservative 8; Mismatches 3; Indels

QY 1 VYVFSTEMANKAAEAVLKQVETIVSFH 28
 IIII: :I IIII: :I : : : :
 Db 238 VYVFTHLANTAAEAVLOGRAESILAYH 265

RESULT 7

ABP06595

ID ABP06595 standard; Protein; 114 AA.

AC ABP06595;

DT 25-JUN-2002 (first entry)

Human ORFX protein sequence SEQ ID NO:13172.

Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KW hypertension; hypothyroidism; cholesterol ester storage disease;
 KW immune deficiency; immune disorder; infectious disease;
 KW autoimmune disorder; autoimmune thyroiditis;
 KW myasthenia gravis.

OS Homo sapiens.

PN WO200192523-A2.

06-DEC-2001.

29-MAY-2001: 2001WO-US10836.

PR 30-MAY-2000: 2000US-206132P.

XX
PR 09-07-0000Z; 0007-Z00S-22870P
XX

PA (CURA-) CURAGEN CORP.
XX

PI SHIMKETS RA, LEACH MD, XX

DR WPI: 2002-106308/14.
DR N-PSDB: ARN22347

PT Novel human polypeptides and polynucleotides useful for diagnosing
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX

PS Disclosure: SEO ID 13172: 1037pp: English.

the present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1 in the specification), ABN15762 to ABN27252 encode the human ORFX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut

CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 114 AA;

Query Match 66.9%; Score 91; DB 23; Length 114;
 Best Local Similarity 57.1%; Pred. No. 1.7e-06;
 Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
 ||||:| || |||||:| :|:|:
 Db 78 VYVFTTHLANTAEEAVLQGRADSILAYH 105

RESULT 8
 AAU78461
 ID AAU78461 standard; Protein; 320 AA.
 XX
 AC AAU78461;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Mouse beta-catenin nuclear localised protein #2.
 XX
 KW Mouse; beta-catenin nuclear localised protein; cancer;
 KW gene therapy; EST; expressed sequence tag.
 XX
 OS Mus musculus.
 XX
 PN WO200224738-A1.
 XX
 PD 28-MAR-2002.
 XX
 PF 19-SEP-2001; 2001WO-JP08140.
 XX
 PR 22-SEP-2000; 2000JP-0287876.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Akiyama T, Adachi S;
 XX
 DR WPI; 2002-330014/36.
 DR N-PSDB; ABK47632.
 XX
 PT New beta-catenin nuclear localised protein for diagnosis and treatment
 PT of diseases associated with nuclear localisation of beta-catenin e.g.
 PT cancer -
 XX
 PS Claim 2; Page 91-92; 113pp; Japanese.
 XX
 CC The invention relates to a beta-catenin nuclear localised protein
 CC and DNA encoding the protein. The protein and encoding DNA are
 CC applicable in diagnosis and treatment of diseases associated with
 CC nuclear localisation of beta-catenin e.g. cancer, including gene
 CC therapy. The present sequence represents the amino acid sequence of
 CC mouse beta-catenin nuclear localised protein #2.
 XX
 SQ Sequence 320 AA;

Query Match 49.3%; Score 67; DB 23; Length 320;
 Best Local Similarity 57.1%; Pred. No. 0.035;
 Matches 12; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 8 MANKAAEAVLKQGVETIVSFH 28
 :|| |||||:| :|:|:
 Db 1 LANTAEEAVLQGRADSILAYH 21

RESULT 9

AAO07544
 ID AAO07544 standard; Protein; 113 AA.
 XX
 AC AAO07544;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 21436.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI87475.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 20; SEQ ID NO 21436; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 113 AA;

Query Match 47.8%; Score 65; DB 22; Length 113;
 Best Local Similarity 75.0%; Pred. No. 0.021;
 Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 8 MANKAAEAVLKQGVETIVSF 27
 | :|||||:| :|:|:
 Db 81 MCVQAAEAVLKQGVETDASF 100

RESULT 10
 AAG33446
 ID AAG33446 standard; Protein; 360 AA.
 XX
 AC AAG33446;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Zea mays protein fragment SEQ ID NO: 40525.

PR	31-AUG-1999;	99US-0151438;
PR	01-SEP-1999;	99US-0151930;
PR	07-SEP-1999;	99US-0152363;
PR	10-SEP-1999;	99US-0153070;
PR	13-SEP-1999;	99US-0153758;
PR	15-SEP-1999;	99US-0154018;
PR	16-SEP-1999;	99US-0154039;
PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-0155139;
PR	23-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157117;
PR	05-OCT-1999;	99US-0157753;
PR	06-OCT-1999;	99US-0157865;
PR	07-OCT-1999;	99US-0158029;
PR	08-OCT-1999;	99US-0158232;
PR	12-OCT-1999;	99US-0158369;
PR	13-OCT-1999;	99US-0159293;
PR	13-OCT-1999;	99US-0159294;
PR	13-OCT-1999;	99US-0159295;
PR	14-OCT-1999;	99US-0159329;
PR	14-OCT-1999;	99US-0159330;
PR	14-OCT-1999;	99US-0159331;
PR	14-OCT-1999;	99US-0159637;
PR	14-OCT-1999;	99US-0159638;
PR	18-OCT-1999;	99US-0159584;
PR	21-OCT-1999;	99US-0160741;
PR	21-OCT-1999;	99US-0160767;
PR	21-OCT-1999;	99US-0160768;
PR	21-OCT-1999;	99US-0160770;
PR	21-OCT-1999;	99US-0160814;
PR	21-OCT-1999;	99US-0160815;
PR	22-OCT-1999;	99US-0160980;
PR	22-OCT-1999;	99US-0160981;
PR	22-OCT-1999;	99US-0160989;
PR	25-OCT-1999;	99US-0161404;
PR	25-OCT-1999;	99US-0161405;
PR	25-OCT-1999;	99US-0161406;
PR	26-OCT-1999;	99US-0161359;
PR	26-OCT-1999;	99US-0161360;
PR	26-OCT-1999;	99US-0161361;
PR	28-OCT-1999;	99US-0161920;
PR	28-OCT-1999;	99US-0161992;
PR	28-OCT-1999;	99US-0161993;
PR	29-OCT-1999;	99US-0162142;

```

Query Match          38.2%;      Score 52;   DB 21;   Length 360;
Best Local Similarity 50.0%;      Pred. NO. 9.4;
Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

```

QY 4 FSTEMANKAAEAVLKGQVETIVSF 27
 ||| :|| ||| ||| :||
 Db 97 FSTRLANNLNVLKGPETIAAF 120

RESULT 11
AAG33445
ID AAG33445 standard; Protein; 448 AA.

AC AAG33445;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 40524.

XX Protein identification; signal transduction

KW hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence; corn.

XX Zea mays subsp. mays.



PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-01343768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135629.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.
PR	21-JUN-1999; 99US-0139817.
PR	22-JUN-1999; 99US-0139899.
PR	23-JUN-1999; 99US-0140353.
PR	23-JUN-1999; 99US-0140354.
PR	24-JUN-1999; 99US-0140695.
PR	28-JUN-1999; 99US-0140823.
PR	29-JUN-1999; 99US-0140991.
PR	30-JUN-1999; 99US-0141287.
PR	01-JUL-1999; 99US-0141842.

PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144632.
PR	21-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.

PR	20-SEP-1999;	99US-0154779;
PR	22-SEP-1999;	99US-0155139;
PR	23-SEP-1999;	99US-0155486;
PR	24-SEP-1999;	99US-0155659;
PR	28-SEP-1999;	99US-0156458;
PR	29-SEP-1999;	99US-0156596;
PR	04-OCT-1999;	99US-0157117;
PR	05-OCT-1999;	99US-0157753;
PR	06-OCT-1999;	99US-0157865;
PR	07-OCT-1999;	99US-0158029;
PR	08-OCT-1999;	99US-0158232;
PR	12-OCT-1999;	99US-0158369;
PR	13-OCT-1999;	99US-0159293;
PR	13-OCT-1999;	99US-0159294;
PR	13-OCT-1999;	99US-0159295;
PR	14-OCT-1999;	99US-0159329;
PR	14-OCT-1999;	99US-0159330;
PR	14-OCT-1999;	99US-0159331;
PR	14-OCT-1999;	99US-0159637;
PR	14-OCT-1999;	99US-0159638;
PR	18-OCT-1999;	99US-0159584;
PR	21-OCT-1999;	99US-0160741;
PR	21-OCT-1999;	99US-0160767;
PR	21-OCT-1999;	99US-0160768;
PR	21-OCT-1999;	99US-0160770;
PR	21-OCT-1999;	99US-0160814;
PR	21-OCT-1999;	99US-0160815;
PR	22-OCT-1999;	99US-0160980;
PR	22-OCT-1999;	99US-0160981;
PR	22-OCT-1999;	99US-0160989;
PR	25-OCT-1999;	99US-0161404;
PR	25-OCT-1999;	99US-0161405;
PR	25-OCT-1999;	99US-0161406;
PR	26-OCT-1999;	99US-0161359;
PR	26-OCT-1999;	99US-0161360;
PR	26-OCT-1999;	99US-0161361;
PR	28-OCT-1999;	99US-0161920;
PR	28-OCT-1999;	99US-0161992;
PR	28-OCT-1999;	99US-0161993;
PR	29-OCT-1999;	99US-0162142;

Query Match 38.2%; Score 52; DB 21; Length 448;

Best Local Similarity 50.0%; Pred. No. 12;

Matches	12;	Conservative	2;	Mismatches	10;	Indels	0;	Gaps	0;
Matches	12;	Conservative	2;	Mismatches	10;	Indels	0;	Gaps	0;

QY 4 FSTEMANKAAEAVLKGQVETIVSF 27
111 : 111 111 : 1
Db 185 FSTRLANLENLVKKEGPETIAAF 208

RESULT 12

AAG33444

ID AAG33444 standard; Protein; 509 AA.

XX
AC

XX
XX

DT 18-OCT-2000 (first entry)

XX
DE
XX
KW
KW
KW

Zea mays protein fragment SEQ ID NO: 40523.

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence; corn.

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 38.2%; Score 52; DB 21; Length 509;

Best Local Similarity 50.0%; Pred. No. 14;

Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQVETIVSF 27
||| :|| ||| ||| :|
Db 246 FSTRLANLENLVKGPETIAAF 269

RESULT 13

ABB60387
ID ABB60387 standard; Protein; 1049 AA.

XX AC ABB60387;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 7953.

XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

DR N-PSDB; ABL04490.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT
XX Disclosure; SEQ ID NO 7953; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1049 AA;

Query Match 37.5%; Score 51; DB 22; Length 1049;

Best Local Similarity 40.7%; Pred. No. 49;

Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQVETIVSFH 28
|| :|| | ||| |::| |
Db 471 YVSCSDMAASATEAVRSGLKIPEHH 497

RESULT 14

ABB89793

ID ABB89793 standard; Protein; 130 AA.

XX AC ABB89793;

XX DT 24-MAY-2002 (first entry)

XX DE Human polypeptide SEQ ID NO 2169.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer; vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder; neurological disease; infection; human; secreted protein.

XX OS Homo sapiens.

XX PN WO200190304-A2.

XX PD 29-NOV-2001.

XX PF 18-MAY-2001; 2001WO-US16450.

XX PR 19-MAY-2000; 2000US-205515P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Birse CE, Rosen CA;

XX DR WPI; 2002-122018/16.

DR N-PSDB; ABL90202.

XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders -

XX Claim 11; SEQ ID NO 2169; 2081pp + Sequence Listing; English.

XX The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB89040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are

quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

```
Query Match      36.8%; Score 50; DB 22; Length 365;
Best Local Similarity 29.6%; Pred. NO. 20;
Matches 8; Conservative 8; Mismatches 11; Indels 0; Gaps 0;
```

Qy 2 YVFSTEMANKAAEAVALKGQVETIVSFH 28
 ||| : | : | : | : | : |
 Db 67 YVFAYPVCNASKATIISGLTECLIHCH 93

Search completed: October 15, 2003, 10:30:22
Job time : 21.214 secs

ID ABG15088 standard; Protein; 365 AA.

AC ABG15088;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #15079.

Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW
KW
food supplement; medical imaging; diagnostic; genetic disorder.
KW

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001

30-MAR-2001: 2001WO-US08631-

31-MAR-2000: 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEO INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 45447; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 5.66537 Seconds
(without alignments)
209.113 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204
Perfect score: 136
Sequence: 1 VYVFSTEMANKAAEAVLKGVETIVSFH 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	34.6	2639	3 US-09-080-983-3	Sequence 3, Appli
2	44	32.4	718	3 US-09-090-808-2	Sequence 2, Appli
3	44	32.4	718	3 US-09-447-453-2	Sequence 2, Appli
4	43	31.6	145	4 US-09-091-725-37	Sequence 37, Appli
5	43	31.6	438	3 US-08-486-099-93	Sequence 93, Appli
6	43	31.6	438	3 US-08-360-107A-103	Sequence 103, App
7	43	31.6	438	3 US-08-484-223B-93	Sequence 93, Appli
8	43	31.6	438	3 US-08-919-597-93	Sequence 93, Appli
9	43	31.6	438	3 US-08-475-668A-93	Sequence 93, Appli
10	43	31.6	438	3 US-08-485-551A-93	Sequence 93, Appli
11	43	31.6	438	3 US-08-471-913A-93	Sequence 93, Appli
12	43	31.6	438	3 US-08-485-264A-93	Sequence 93, Appli
13	43	31.6	438	4 US-08-474-349A-93	Sequence 93, Appli
14	43	31.6	438	4 US-08-255-208A-29	Sequence 29, Appli
15	43	31.6	438	4 US-08-470-896-93	Sequence 93, Appli
16	43	31.6	438	4 US-08-485-546A-93	Sequence 93, Appli
17	43	31.6	662	1 US-08-224-657-88	Sequence 88, Appli
18	43	31.6	662	4 US-09-354-138-88	Sequence 88, Appli
19	42	30.9	154	4 US-09-732-210-103	Sequence 103, App
20	42	30.9	154	4 US-09-732-210-210	Sequence 210, App
21	42	30.9	188	4 US-09-252-991A-29737	Sequence 29737, A
22	42	30.9	300	4 US-09-634-238-219	Sequence 219, App
23	42	30.9	358	4 US-09-934-901-8	Sequence 8, Appli
24	42	30.9	541	4 US-09-627-376-11	Sequence 11, Appli
25	42	30.9	1820	3 US-07-998-289B-8	Sequence 11, Appli
26	41.5	30.5	111	4 US-09-775-932-20	Sequence 20, Appli
27	41.5	30.5	678	1 US-08-844-085-2	Sequence 2, Appli

28 41.5 30.5 745 1 US-08-453-472-5 Sequence 5, Appli
29 41.5 30.5 745 1 US-08-038-948-9 Sequence 9, Appli
30 41.5 30.5 745 1 US-08-453-952-5 Sequence 5, Appli
31 41.5 30.5 745 2 US-08-484-993B-43 Sequence 43, Appli
32 41.5 30.5 745 2 US-08-862-903-5 Sequence 5, Appli
33 41.5 30.5 745 2 US-08-484-158B-43 Sequence 43, Appli
34 41.5 30.5 745 2 US-08-484-596A-43 Sequence 43, Appli
35 41.5 30.5 745 2 US-08-480-150A-43 Sequence 43, Appli
36 41.5 30.5 745 3 US-08-458-731-43 Sequence 43, Appli
37 41.5 30.5 745 3 US-08-149-223A-43 Sequence 43, Appli
38 41 30.1 111 4 US-09-107-532A-5449 Sequence 5449, Ap
39 41 30.1 162 4 US-09-134-001C-3670 Sequence 3670, Ap
40 41 30.1 211 3 US-09-075-194-1 Sequence 1, Appli
41 41 30.1 256 2 US-08-578-516-6 Sequence 6, Appli
42 41 30.1 378 4 US-09-107-532A-6500 Sequence 6500, Ap
43 41 30.1 569 4 US-09-107-532A-6689 Sequence 6689, Ap
44 41 30.1 676 4 US-09-107-532A-5806 Sequence 5806, Ap
45 41 30.1 689 4 US-09-499-964-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-080-983-3
; Sequence 3, Application US/09080983
; Patent No. 6197948
; GENERAL INFORMATION:
; APPLICANT: Zhu, Hai-Ying
; APPLICANT: Ling, Kai-Shu
; APPLICANT: Gonsalves, Dennis
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS TYPE 2 PROTEINS
; TITLE OF INVENTION: AND THEIR USES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,983
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/047,194
; FILING DATE: 20-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1631
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2639 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-080-983-3

Query Match 34.6%; Score 47; DB 3; Length 2639;
Best Local Similarity 33.3%; Pred. No. 1.3e+02;
Matches 9; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKGVETIVSFH 28
:||||: :| | | | : ||
Db 1788 HLFSSDDVDDSSASAGLKGASRMTLFH 1814

RESULT 2

US-09-090-808-2
; Sequence 2, Application US/09090808
; Patent No. 6127159
; GENERAL INFORMATION:
; APPLICANT: Fuller, Margaret
; APPLICANT: Hales, Karen
; TITLE OF INVENTION: Mitofusin Genes and their
; NUMBER OF INVENTION: Uses
; CORRESPONDENCE ADDRESSES: 6
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/090,808
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/048,961
; FILING DATE: 06-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SUN-63P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650 327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-090-808-2

Query Match 32.4%; Score 44; DB 3; Length 718;
Best Local Similarity 34.6%; Pred. No. 79;
Matches 9; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 2 YVFS--TEMANKAAEAVLKGVETIV 25
| | | | : | | | : ||
Db 403 YCFEELTEMTQVRGRCVLSQIKTLI 428

RESULT 3

US-09-447-453-2
; Sequence 2, Application US/09447453
; Patent No. 6284507
; GENERAL INFORMATION:
; APPLICANT: Fuller, Margaret
; APPLICANT: Hales, Karen
; TITLE OF INVENTION: Mitofusin Genes and their
; NUMBER OF INVENTION: Uses
; CORRESPONDENCE ADDRESSES: 6
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Avenue, Suite 200

; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/447,453
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 06/048,961
; FILING DATE: 06-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sherwood, Pamela J
; REGISTRATION NUMBER: 36,677
; REFERENCE/DOCKET NUMBER: SUN-63P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-327-3400
; TELEFAX: 650 327-3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-447-453-2

Query Match 32.4%; Score 44; DB 3; Length 718;
Best Local Similarity 34.6%; Pred. No. 79;
Matches 9; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 2 YVFS--TEMANKAAEAVLKGVETIV 25
| | | | : | | | : ||
Db 403 YCFEELTEMTQVRGRCVLSQIKTLI 428

RESULT 4

US-09-091-725-37
; Sequence 37, Application US/09091725
; Patent No. 6329141
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Improved methods for transforming Phaffia
; and recombinant DNA for use therein
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster llp
; STREET: 2000 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: United States of America
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,725
; FILING DATE: 23-DEC-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95203620.0
; FILING DATE: 22-DEC-1995
; APPLICATION NUMBER: EP 96200943.7
; FILING DATE: 11-APR-1996
; ATTORNEY/AGENT INFORMATION:


```
; NAME: E. Victor Donahue
; REGISTRATION NUMBER: 35,492
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-091-725-37

Query Match 31.6%; Score 43; DB 4; Length 145;
Best Local Similarity 43.8%; Pred. No. 16;
Matches 14; Conservative 4; Mismatches 6; Indels 8; Gaps 2;

QY 5 STEMANKAA-EAVLKG-----QVETIVSFH 28
| : : | | | | : | | | | |
Db 15 SADVKAAKAAKAAKGTQSTSTRKVRTSVSFH 46

RESULT 5
US-08-486-099-93
; Sequence 93, Application US/08486099
; Patent No. 6013263
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,099
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-031
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-486-099-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKGQVETIVS 26
| : : | | | | : | | | | |

US-08-484-223B-93
; Sequence 93, Application US/08484223B
; Patent No. 6020459
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
```

```
Db 41 SLEQSNKAIEIREATQETVIA 62
| : : | | | | : | | | | |

RESULT 6
US-08-360-107A-103
; Sequence 103, Application US/08360107A
; Patent No. 6017536
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; NUMBER OF SEQUENCES: 149
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,107A
; FILING DATE: 20-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 103:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-360-107A-103

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKGQVETIVS 26
| : : | | | | : | | | | |

Db 41 SLEQSNKAIEIREATQETVIA 62
| : : | | | | : | | | | |

RESULT 7
US-08-484-223B-93
; Sequence 93, Application US/08484223B
; Patent No. 6020459
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
```



```
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 245
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,223B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-484-223B-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : | | | | : | | : |
Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 8
US-08-919-597-93
; Sequence 93, Application US/08919597
; Patent No. 6054265
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,597
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470,896
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amido acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-919-597-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : | | | | : | | : |
Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 9
US-08-475-668A-93
; Sequence 93, Application US/08475668A
; Patent No. 6060065
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,668A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
```

; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-475-668A-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : || | : : ||::
Db 41 SLEQSNKAIEEIREATQETVIA 62

RESULT 10
US-08-485-551A-93
; Sequence 93, Application US/08485551A
; Patent No. 6068973
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 211
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,551A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-023
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-485-551A-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : || | : : ||::
Db 41 SLEQSNKAIEEIREATQETVIA 62

RESULT 11

US-08-471-913A-93
; Sequence 93, Application US/08471913A
; Patent No. 6093794
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 214
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,913A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-471-913A-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : || | : : ||::
Db 41 SLEQSNKAIEEIREATQETVIA 62

RESULT 12

US-08-485-264A-93
; Sequence 93, Application US/08485264A
; Patent No. 6228983
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING

;
; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,264A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-264A-93

Query Match 31.6%; Score 43; DB 3; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQVETIVS 26
| | : | | | : : : | | : :
Db 41 SLEQSNKAIEEIREATQETVIA 62

RESULT 13
US-08-474-349A-93
; Sequence 93, Application US/08474349A
; Patent No. 6333395
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; TITLE OF INVENTION: VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

;
; APPLICATION NUMBER: US/08/474,349A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-474-349A-93

Query Match 31.6%; Score 43; DB 4; Length 438;
Best Local Similarity 36.4%; Pred. No. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQVETIVS 26
| | : | | | : : : | | : :
Db 41 SLEQSNKAIEEIREATQETVIA 62

RESULT 14
US-08-255-208A-29
; Sequence 29, Application US/08255208A
; Patent No. 6440656
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway Jr., Stephen R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/255,208A
; FILING DATE: 07-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

US-08-255-208A-29

Query Match 31.6%; Score 43; DB 4; Length 438;
Best Local Similarity 36.4%; Pred. NO. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : | | | | : : | | : :
Db 41 SLEQSNKAIEIREATQETVIA 62

RESULT 15
US-08-470-896-93
; Sequence 93, Application US/08470896
; Patent No. 6479055
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway, Stephen R.
; APPLICANT: Langlois, Alphonse J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,896
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-470-896-93

Query Match 31.6%; Score 43; DB 4; Length 438;
Best Local Similarity 36.4%; Pred. NO. 63;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKQGVETIVS 26
| | : | | | | : : | | : :
Db 41 SLEQSNKAIEIREATQETVIA 62

Search completed: October 15, 2003, 10:31:58
Job time : 6.66537 secs


```
; Sequence 15, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/322,579
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US/09/915,543
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,502
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 1426
; TYPE: PRT
; ORGANISM: Human lgs/bcl9
US-10-322-579-15

Query Match      100.0%; Score 136; DB 15; Length 1426;
Best Local Similarity 100.0%; Pred. No. 9.4e-13;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
Db      177 VYVFSTEMANKAAEAVLKQGVETIVSFH 204

RESULT 3
US-10-322-579-2
; Sequence 2, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/322,579
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US/09/915,543
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,502
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Drosophila lgs
US-10-322-579-2

Query Match      72.1%; Score 98; DB 15; Length 28;
Best Local Similarity 57.1%; Pred. No. 9.6e-09;
Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY      1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28
Db      1 IFVFSTQLANKGAESVLSGQFQTIIAYH 28

RESULT 4
US-10-128-714-3547
; Sequence 3547, Application US/10128714
```

```
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3547
; LENGTH: 984
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3547
```

```
Query Match      36.0%; Score 49; DB 15; Length 984;
Best Local Similarity 52.6%; Pred. No. 46;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      3 VFSTEMANKAAEAVLKGV 21
Db      500 VYRANMANKSAAAVLKSKL 518
```

```
RESULT 5
US-10-128-714-8547
; Sequence 8547, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8547
; LENGTH: 1058
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8547
```


Query Match 36.0%; Score 49; DB 15; Length 1058;
Best Local Similarity 52.6%; Pred. No. 50;
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 VFSTEMANKAAEAVLKGV 21
| : ||||| : ||||| :
Db 500 VYRANMANKSAAAVLKSKL 518

RESULT 6
US-10-008-355-6
; Sequence 6, Application US/10008355
; Publication No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan s
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 732
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens

US-10-008-355-6
Query Match 34.6%; Score 47; DB 14; Length 732;
Best Local Similarity 39.1%; Pred. No. 67;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKGVET 23
| : | | | : | | : | : |
Db 517 VALYDTNMAQEKAKILAGKLST 539

RESULT 7
US-10-167-547C-8
; Sequence 8, Application US/10167547C
; Publication No. US20030170653A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; APPLICANT: Damude, Howard G.
; TITLE OF INVENTION: A Biological Method for the Production of Alpha-Methylene-Gamma
; FILE REFERENCE: CL1804 US NA
; CURRENT APPLICATION NUMBER: US/10/167,547C
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: 60/297198
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Microsoft Office 07
; SEQ ID NO 8
; LENGTH: 471
; TYPE: PRT
; ORGANISM: tulip pistil

US-10-167-547C-8
Query Match 33.8%; Score 46; DB 12; Length 471;
Best Local Similarity 41.7%; Pred. No. 57;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKGVETIVSF 27
| | : | | : : | | | : |
Db 204 FSTRLANNLEKILTEGPETIAAF 227

RESULT 8

US-10-167-547C-10
; Sequence 10, Application US/10167547C
; Publication No. US20030170653A1
; GENERAL INFORMATION:
; APPLICANT: E.I. du Pont de Nemours and Company
; APPLICANT: Damude, Howard G.
; TITLE OF INVENTION: A Biological Method for the Production of Alpha-Methylene-Gam
; FILE REFERENCE: CL1804 US NA
; CURRENT APPLICATION NUMBER: US/10/167,547C
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: 60/297198
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Microsoft Office 07
; SEQ ID NO 10
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Alstroemeria

US-10-167-547C-10
Query Match 33.1%; Score 45; DB 12; Length 507;
Best Local Similarity 37.5%; Pred. No. 90;
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKGVETIVSF 27
| | | : | | : | : | : | : |
Db 246 FSTRLANNLELIVKEGPDITAAF 269

RESULT 9
US-10-156-761-10721
; Sequence 10721, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10721
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis

US-10-156-761-10721
Query Match 33.1%; Score 45; DB 15; Length 747;
Best Local Similarity 39.1%; Pred. No. 1.4e+02;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 3 VFSTEMANKAAEAVLKGVETIV 25
| | : | | : | : | : | : |
Db 250 VFATSSLNRARQALDRGTERAV 272

RESULT 10
US-10-342-224-92
; Sequence 92, Application US/10342224
; Publication No. US20030162294A1
; GENERAL INFORMATION:
; APPLICANT: Nathalie Verbruggen
; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress


```

; FILE REFERENCE: CNN-012US
; CURRENT APPLICATION NUMBER: US/10/342,224
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US/09/762,154
; PRIOR FILING DATE: 2002-02-02
; PRIOR APPLICATION NUMBER: EP 98202634.6
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-342-224-92

```

Query Match	32.7%	Score 44.5;	DB 12;	Length 170;
Best Local Similarity	52.0%;	Pred. NO. 29;		
Matches 13; Conservative	0;	Mismatches 7;	Indels 5;	Gaps 1;

Qy	6	TEMANKAAEAVL-----KGQVETIV	25
Db	85	TEYVRKTVVEVVLTDLTLEKKQVETIV	109

RESULT 11

```

US-10-156-761-9002
; Sequence 9002, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9002
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9002

```

Query Match 32.4%; Score 44; DB 15; Length 171;
Best Local Similarity 35.7%; Pred. No. 35;
Matches 10; Conservative 5; Mismatches 13; Indels 0;
Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28
Db 29 IFCYRASMAAKKAEQVLYDOWRGILAVH 56

RESULT 12

US-09-815-242-10457
; Sequence 10457, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```

; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10457
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10457

```

Query Match 32.4%; Score 44; DB 9; Length 484;
Best Local Similarity 30.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 7; Mismatches 7. Indels

QY 8 MANKAAEAVLKQVETIVSF 27
:
Db 52 LLRKSVSVIIKGTIKTIIGF 71

RESULT 13

```

US-10-117-846-2
; Sequence 2, Application US/10117846
; Publication No. US20020168673A1
; GENERAL INFORMATION:
; APPLICANT: Fuller, Margaret T
; APPLICANT: Hales, Karen G.
; APPLICANT: Santel, Ansgar H.
; TITLE OF INVENTION: Mitofusins, Fzo Homologs and Functional
; TITLE OF INVENTION: Derivatives Thereof
; FILE REFERENCE: STAN-063CIP3
; CURRENT APPLICATION NUMBER: US/10/117,846
; CURRENT FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: 09/413,285
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: PCT/US00/27871
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 718
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-117-846-2

```

Query Match

Query Match 32.4%; Score 44; DB 14; Length 718;
Best Local Similarity 34.6%; Pred. No. 2e+02;
Matches 9; Conservative 5; Mismatches 10; Indels

QY 2 YVFS--TEMANKAAEAVLKQVETIV 25
+ + + + + : + + + + +
Db 403 YCFEELTEMTQVGRCVLSDOIKTLL 428

RESULT 14

US-09-764-891-5144
; Sequence 5144, Application US/09764891

```
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5144
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-5144

Query Match      31.6%; Score 43; DB 11; Length 80;
Best Local Similarity 34.8%; Pred. No. 20;
Matches 8; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQVET 23
Db 46 MHFSSEYANKPVTASAMHMQT 68
:::|:| | | | :
:

RESULT 15
US-10-306-762-212
; Sequence 212, Application US/10306762
; Publication No. US20030187220A1
; GENERAL INFORMATION:
; APPLICANT: Park, Frances
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF A FLAVIN
; FILE REFERENCE: 52498-20011.00
; CURRENT APPLICATION NUMBER: US/10/306,762
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/334,132
; PRIOR FILING DATE: 2001-11-28
; NUMBER OF SEQ ID NOS: 242
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 212
; LENGTH: 315
; TYPE: PRT
; ORGANISM: A. tumefaciens (15888840)
US-10-306-762-212

Query Match      31.6%; Score 43; DB 12; Length 315;
Best Local Similarity 39.1%; Pred. No. 1.1e+02;
Matches 9; Conservative 6; Mismatches 6; Indels 2; Gaps 1;

QY 6 TEMANKAAEAVLKQVETIVSFH 28
Db 29 TEMI--VADAIHQREKLLGYH 49
| | | | : | : | | : |
:

Search completed: October 15, 2003, 10:50:57
Job time : 13.8755 secs
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 3.37743 Seconds
(without alignments)
389.867 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204
Perfect score: 136
Sequence: 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	1426	1 BCL9_HUMAN	O00512 homo sapien
2	98	72.1	1469	1 BCL9_DROME	Q961d9 drosophila
3	60	44.1	205	1 HEM2_CLOJO	Q59295 clostridium
4	51	37.5	1217	1 SYV_FUGRU	P49696 fugu rubrip
5	46	33.8	330	1 QUTG_EMENI	P25416 emericeella
6	46	33.8	363	1 BUK_BACSU	P54532 bacillus su
7	46	33.8	2273	1 HFAL_YEAST	P32874 saccharomyc
8	45.5	33.5	224	1 SODM_CHAFE	O96347 charybdis f
9	45.5	33.5	444	1 GUNN_ERWCA	Q59394 erwinia car
10	45.5	33.5	504	1 GUNW_ERWCA	Q59395 erwinia car
11	45.5	33.5	505	1 GUNW_ERWCA	Q47096 erwinia car
12	45.5	33.5	745	1 ZP2_MACRA	O77726 macaca radi
13	45	33.1	490	1 PEM3_ARATH	Q9c6b9 arabidopsis
14	44.5	32.7	170	1 AIG2_ARATH	P54121 arabidopsis
15	44	32.4	318	1 YB64_METJA	Q58564 methanococc
16	44	32.4	340	1 MPCP_CAEEL	P40614 caenorhabdi
17	44	32.4	484	1 SGAT_ECOLI	P39301 escherichia
18	44	32.4	602	1 DNLJ_AERPE	Q9yd18 aeropyrum p
19	44	32.4	629	1 SYR_PYRFU	O8u149 pyrococcus
20	44	32.4	980	1 SYV_SCHPO	O75005 schizosacch
21	44	32.4	4427	1 PKSL_BACSU	O05470 bacillus su
22	43	31.6	119	1 TH12_CAEEL	Q17424 caenorhabdi
23	43	31.6	159	1 RIB4_SCHPO	Q9uub1 schizosacch
24	43	31.6	291	1 EX53_MYCPN	P75403 mycoplasma
25	43	31.6	345	1 TRPD_SULTO	Q97127 sulfolobus
26	43	31.6	537	1 AREH_SCHPO	Q10269 schizosacch
27	43	31.6	575	1 YDGE_SCHPO	Q10499 schizosacch
28	43	31.6	594	1 SYV_RAT	Q04462 rattus norv
29	43	31.6	602	1 GLSL_HUMAN	Q9ui32 homo sapien
30	43	31.6	609	1 NPRV_VIBPR	Q00971 vibrio prot
31	43	31.6	611	1 EMPA_VIBAN	P43147 vibrio angu
32	43	31.6	662	1 VGLF_CDVO	P12569 canine dist
33	43	31.6	895	1 SECA_CYACA	O19911 cyanidium c

RESULT 1					
BCL9_HUMAN					
ID	BCL9_HUMAN	STANDARD;	PRT;	1426	AA.
AC	O00512;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).				
GN	BCL9.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Fetal brain;				
RX	MEDLINE=98158621; PubMed=9490669;				
RA	Willis T.G., Zalcborg I.R., Coignet L.J.A., Wlodarska I., Stul M.,				
RA	Jadavel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,				
RA	Dyer M.J.S.;				
RT	"Molecular cloning of translocation t(1;14)(q21;q32) defines a novel				
RT	gene (BCL9) at chromosome 1q21.";				
RL	Blood 91:1873-1881(1998).				
RN	[2]				
RP	FUNCTION.				
RX	MEDLINE=21952490; PubMed=11955446;				
RA	Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,				
RA	Murone M., Zuellig S., Basler K.;				
RT	"Wnt/wingless signaling requires BCL9/legless-mediated recruitment of				
RT	pygopus to the nuclear beta-catenin-TCF complex.";				
RL	Cell 109:47-60(2002).				
CC	-!- FUNCTION: Involved in signal transduction through the wnt pathway.				
CC	-!- SUBUNIT: Binds to beta-catenin (CTNBN1), PYGO1 and PYGO2.				
CC	-!- SUBCELLULAR LOCATION: Nuclear (Probable).				
CC	-!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,				
CC	testis, ovary and small intestine, and at lower levels in spleen,				
CC	colon and blood.				
CC	-!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation				
CC	found in a patient with precursor B-cell acute lymphoblastic				
CC	leukemia (ALL). This translocation leaves the coding region				
CC	intact, but may have pathogenic effects due to alterations in the				
CC	expression level of BCL9. Several cases of translocations within				
CC	the 3' untranslated region of BCL9 have been found in B-cell				
CC	malignancies.				
CC	-!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.				
CC	-!- CAUTION: Ref.1 sequence differs from that shown due to a				
CC	frameshift in position 1391.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				

Q15269 homo sapien
P30733 solanum tub
Q9z1q9 mus musculu
P26640 homo sapien
P28367 bacillus su
P93407 oryza sativ
Q8ty76 methanopyru
P70955 bacillus su
Q57672 methanococc
Q9vhw7 drosophila
O67887 aquifex aeo
O94065 candida alb

ALIGNMENTS


```
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (Porphobilinogen
DE synthase) (ALAD) (ALADH) (Fragment).
GN HEMB.
OS Clostridium josui.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1499;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FERM P-9684; PubMed=7665501;
RX Fujino E., Fujino T., Karita S., Sakka K., Ohmiya K.;
RT "Cloning and sequencing of some genes responsible for porphyrin
RT biosynthesis from the anaerobic bacterium Clostridium josui.";
RL J. Bacteriol. 177:5169-5175(1995).
CC -!- CATALYTIC ACTIVITY: 2 5-aminolevulinate = porphobilinogen + 2
CC H(2)O.
CC -!- COFACTOR: Zinc (By similarity).
CC -!- PATHWAY: Siroheme biosynthesis.
CC -!- SUBUNIT: Homooctamer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE ALADH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D28503; BAA05863.1; -.
DR PIR; I40812; I40812.
DR HSSP; P15002; 1B4E.
DR InterPro; IPR001731; AlaD_dehydratase.
DR Pfam; PF00490; ALAD; 1.
DR PRINTS; PR00144; DALDHYDRTASE.
DR ProDom; PD002304; AlaD_dehydratase; 1.
DR PROSITE; PS00169; D_ALA_DEHYDRATASE; PARTIAL.
KW Porphyrin biosynthesis; Lyase; Zinc.
FT DOMAIN 114 132 ZINC-BINDING (BY SIMILARITY).
FT NON_TER 205 205
SQ SEQUENCE 205 AA; 23172 MW; 886F9DAEFD81144E CRC64;

Query Match 44.1%; Score 60; DB 1; Length 205;
Best Local Similarity 46.2%; Pred. No. 0.1;
Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQGVETIVSF 27
| | | : | | | | | | | | | | | | |
Db 51 YHFSPDMVGKATAEAAALKADVKSVLLF 76

RESULT 4
SYV_FUGRU
ID SYV_FUGRU STANDARD; PRT; 1217 AA.
AC P49696;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).
GN VARS1.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97396021; PubMed=9254008;
```

```
RA Lim E.H., Corrochano L.M., Elgar G., Brenner S.;
RT "Genomic structure and sequence analysis of the valyl-tRNA synthetase
RT gene of the Japanese pufferfish, Fugu rubripes.";
RL DNA Seq. 7:141-151(1997).
CC -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
CC + L-valyl-tRNA(Val).
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -!- SIMILARITY: THE N-TERMINAL DOMAIN IS SIMILAR TO ELONGATION
CC FACTOR 1-GAMMA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91856; CAA62967.1; -.
DR HSSP; P96142; 1GAX.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002303; tRNA-synt_val.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00986; TRNASYNTHVAL.
DR TIGRFAMS; TIGR00422; vals; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT DOMAIN 1 7320 EF-1-GAMMA LIKE.
FT SITE 293 303 "HIGH" REGION.
FT SITE 809 813 "KMSKS" REGION.
FT BINDING 812 812 ATP (BY SIMILARITY).
SQ SEQUENCE 1217 AA; 138218 MW; 5E08AF24B5C8A7A1 CRC64;

Query Match 37.5%; Score 51; DB 1; Length 1217;
Best Local Similarity 37.0%; Pred. No. 14;
Matches 10; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQGVETIVSFH 28
| | : | | : | | : | | : | | : | |
Db 626 YVSCSDMGKQAADAVREGRLKIIPDHH 652

RESULT 5
QUTG_EMENI
ID QUTG_EMENI STANDARD; PRT; 330 AA.
AC P25416;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE QUTG protein.
GN QUTG.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89181521; PubMed=2976880;
RA Hawkins A.R., Lamb H.K., Smith M., Keyte J.W., Roberts C.F.;
RT "Molecular organisation of the quinic acid utilization (QUT) gene
RT cluster in Aspergillus nidulans.";
RL Mol. Gen. Genet. 214:224-231(1988).
RN [2]
RP SIMILARITY TO INOSITOL MONOPHOSPHATASE.
RX MEDLINE=91080861; PubMed=2175387;
RA Lamb H.K., Hawkins A.R., Smith M., Harvey I.J., Brown J., Turner G.,
RA Roberts C.F.;
RT "Spatial and biological characterisation of the complete quinic acid
RT utilisation gene cluster in Aspergillus nidulans.";
RL Mol. Gen. Genet. 223:17-23(1990).
```



```
RL Nature 387:90-93(1997).
RN [2]
RP SEQUENCE OF 125-949 FROM N.A.
RX MEDLINE=941146412; PubMed=7906156;
RA Kearsey S.E.;
RT *Identification of a Saccharomyces cerevisiae gene closely related to
RL FAS3 (acetyl-CoA carboxylase).";
RL DNA Seq. 4:69-70(1993).
CC -|- COFACTOR: BIOTIN (BY SIMILARITY).
CC -|- SIMILARITY: STRONG, TO ACETYL-COA CARBOXYLASE.
CC -|- CAUTION: THE READING FRAME FROM WHICH THIS PROTEIN IN TRANSLATED
CC HAS NO MET INITIATION CODON NEAR TO THE 5'END. IT DOES NOT SEEM TO
CC BE A PSEUDOGENE. THERE ARE NO APPARENT FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z49809; CAA89922.1; -.
CC EMBL; Z48755; CAA88647.1; -.
CC EMBL; Z22558; CAA80280.1; -.
CC PIR; S55089; S55089.
CC HSP; P24182; 1DV1.
CC SGD; S0004820; HFA1.
CC InterPro; IPR001882; Biotin_attach.
CC InterPro; IPR005482; Biotin_carb_C.
CC InterPro; IPR000089; Biotin_lipoyl.
CC InterPro; IPR000022; Carboxyl_trans.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005481; CPase_L_N.
CC Pfam; PF02785; Biotin_carb_C; 1.
CC Pfam; PF00364; biotin_lipoyl; 1.
CC Pfam; PF01039; Carboxyl_trans; 1.
CC Pfam; PF00289; CPSase_L_chain; 1.
CC Pfam; PF02786; CPSase_L_D2; 1.
CC PROSITE; PS00188; BIOTIN; 1.
CC PROSITE; PS00866; CPSASE_1; 1.
CC PROSITE; PS00867; CPSASE_2; 1.
CC Biotin; Ligase; ATP-binding.
KW NP_BIND 332 337 ATP (BY SIMILARITY).
FT ACT_SITE 459 459 BY SIMILARITY.
FT BINDING 804 804 BIOTIN (BY SIMILARITY).
FT CONFLICT 661 661 F -> L (IN REF. 2).
SQ SEQUENCE 2273 AA; 259160 MW; 08727A301549DA92 CRC64;

Query Match 33.8%; Score 46; DB 1; Length 2273;
Best Local Similarity 45.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAEAVLKGVQV 21
Db 659 YVFTEKVRNKYLELLRRGQV 678

RESULT 8
SODM_CHAFE STANDARD; PRT; 224 AA.
AC O96347;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Superoxide dismutase [Mn], mitochondrial precursor (EC 1.15.1.1).
OS Charybdis feriatius (Crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
OC Eubranchyura; Portunoidae; Portunidae; Charybdis.
OX NCBI_TaxID=65693;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RA Lin C.T., Lai Y.S., Kuo T.J., Chang T.C.;
RT "Molecular cloning, expression, and characterization of a cDNA
RT encoding Mn-superoxide dismutase from crab Charybdis feriatius.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: Destroys radicals which are normally produced within the
CC cells and which are toxic to biological systems.
CC -|- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).
CC -|- COFACTOR: Manganese (By similarity).
CC -|- SUBUNIT: Homotetramer (By similarity).
CC -|- SUBCELLULAR LOCATION: Mitochondrial matrix.
CC -|- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF019411; AAD01640.1; -.
CC HSP; P04179; 1ABM.
CC InterPro; IPR001189; SODismutase.
CC Pfam; PF00081; sodfe; 1.
CC Pfam; PF02777; sodfe_C; 1.
CC ProDom; PD000475; SODismutase; 1.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Manganese; Mitochondrion; Transit peptide.
FT TRANSIT 1 20 MITOCHONDRION (BY SIMILARITY).
FT CHAIN 21 224 SUPEROXIDE DISMUTASE [MN].
FT METAL 46 46 MANGANESE (BY SIMILARITY).
FT METAL 94 94 MANGANESE (BY SIMILARITY).
FT METAL 177 177 MANGANESE (BY SIMILARITY).
FT METAL 181 181 MANGANESE (BY SIMILARITY).
SQ SEQUENCE 224 AA; 24527 MW; 107CF19382E9138A CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 224;
Best Local Similarity 50.0%; Pred. No. 17;
Matches 13; Conservative 3; Mismatches 9; Indels 1; Gaps 1;

QY 2 YVFSTEMA-NKAAEAVLKGVETIVS 26
Db 54 YVNNLNVAEEKLAEEKEKGDVSTIIS 79

RESULT 9
GUNN_ERWCA STANDARD; PRT; 444 AA.
AC Q59394;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)
DE (Cellulase N).
GN CELN.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Atroseptica FCBR C18;
RX MEDLINE=98299944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,
RA von Wettstein D.;
RT "Transplanting two unique beta-glucanase catalytic activities into
RT one multienzyme, which forms glucose.";
RL Biotechnology 14:71-76(1996).
CC -|- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
```

CC HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L39788; AAC37033.1; -.

DR HSSP; O85465; 1A3H.

DR InterPro; IPR001956; CBD_3.

DR InterPro; IPR001547; Glyco_hydro_5.

DR Pfam; PF00942; CBM_3; 1.

DR Pfam; PF00150; cellulase; 1.

DR PRODom; PD001947; CBD_3; 1.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 444 ENDOGLUCANASE N.

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 444 AA; 48300 MW; FA7E4179004CBB43 CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 444;

Best Local Similarity 36.7%; Pred. No. 33;

Matches 11; Conservative 5; Mismatches 11; Indels 3; Gaps 1;

QY 2 YVFSTEMANKAAEAFLKQ---VETIVSFH 28

1: : :||| ||| | | | :|

Db 101 YISNPSLANKVKEAVAAQAQSLGVYIIIDWH 130

RESULT 10

GUNW_ERWCA

ID GUNW_ERWCA STANDARD; PRT; 504 AA.

AC Q59395;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Endoglucanase V1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)

DE (Cellulase V1).

GN CELV1.

OS Erwinia carotovora.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pectobacterium.

OX NCBI_TaxID=554;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SCC3193;

RX MEDLINE=95231512; PubMed=7715600;

RA Mae A.; Heikinheimo R.; Palva E.T.;

RT "Structure and regulation of the Erwinia carotovora subspecies

RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in

RT phytopathogenicity.";

RL Mol. Gen. Genet. 247:17-26(1995).

CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

CC linkages in cellulose, lichenin and cereal beta-D-glucans.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL

CC HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X79241; CAA55823.1; -.

DR PIR; S54744; S54744.

DR HSSP; O85465; 1A3H.

DR InterPro; IPR001956; CBD_3.

DR InterPro; IPR001547; Glyco_hydro_5.

DR Pfam; PF00942; CBM_3; 1.

DR Pfam; PF00150; cellulase; 1.

DR PRODom; PD001947; CBD_3; 1.

DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 504 ENDOGLUCANASE V1.

FT DOMAIN 32 334 CATALYTIC.

FT DOMAIN 335 352 LINKER.

FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

SQ SEQUENCE 504 AA; 54963 MW; OD7ECF74781565FA CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 504;

Best Local Similarity 36.7%; Pred. No. 38;

Matches 11; Conservative 5; Mismatches 11; Indels 3; Gaps 1;

QY 2 YVFSTEMANKAAEAFLKQ---VETIVSFH 28

1: : :||| ||| | | | :|

Db 101 YIANPSLANKVKEAVAAQAQSLGVYIIIDWH 130

RESULT 11

GUNV_ERWCA

ID GUNV_ERWCA STANDARD; PRT; 505 AA.

AC Q47096;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)

DE (Cellulase V).

GN CELV.

OS Erwinia carotovora.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Pectobacterium.

OX NCBI_TaxID=554;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SCRI193;

RX MEDLINE=94067016; PubMed=8246888;

RA Cooper V.J.C.; Salmond G.P.C.;

RT "Molecular analysis of the major cellulase (CelV) of Erwinia

RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme

RT domains.";

RL Mol. Gen. Genet. 241:341-350(1993).

CC -!- FUNCTION: Endoglucanase with some exoglucanase activity.

CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic

CC linkages in cellulose, lichenin and cereal beta-D-glucans.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: Has a pH optimum of about 7.0 and a temperature

CC optimum about 42 degrees Celsius.

CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL

CC HYDROLASES).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X76000; CAA53592.1; -.

DR PIR; S39962; S39962.

DR HSSP; O85465; 1A3H.

DR InterPro; IPR001956; CBD_3.

DR InterPro; IPR001547; Glyco_hydro_5.

DR Pfam; PF00942; CBM_3; 1.

DR Pfam; PF00150; cellulase; 1.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 15, 2003, 10:27:33 ; Search time 16.2335 Seconds
(without alignments)
445.097 Million cell updates/sec

Title: US-09-915-543-15_COPY_177_204
Perfect score: 136
Sequence: 1 VYVFSTEMANKAAEAVLKQVETIVSFH 28

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL_23:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriap:*
 - 17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	55	40.4	1050	5	Q9U1Q4	Q9ulq4 caenorhabdi
2	51	37.5	243	16	Q97JA0	Q97ja0 clostridium
3	51	37.5	1049	5	Q9V6L1	Q9v6l1 drosophila
4	51	37.5	1049	5	Q960E6	Q960e6 drosophila
5	50	36.8	459	10	O82521	O82521 capsicum ch
6	50	36.8	642	4	Q9H6R2	Q9h6r2 homo sapien
7	50	36.8	642	4	Q96GN2	Q96gn2 homo sapien
8	50	36.8	1098	4	Q96Q02	Q96q02 homo sapien
9	49.5	36.4	350	2	Q8GGI9	Q8ggi9 lactobacill
10	49	36.0	141	10	Q9LFW7	Q9lfw7 arabidopsis
11	49	36.0	143	17	Q976C9	Q976c9 sulfolobus
12	49	36.0	187	16	Q8G3S3	Q8g3s3 bifidobacte
13	49	36.0	210	10	Q9FZL1	Q9fzl1 arabidopsis
14	49	36.0	233	17	Q96YA9	Q96ya9 sulfolobus
15	49	36.0	237	17	Q96X55	Q96x55 sulfolobus
16	49	36.0	268	16	Q8DFX2	Q8dfx2 vibrio vuln

17	49	36.0	458	16	O83532	O83532 treponema p
18	49	36.0	586	16	Q9KMS4	Q9kms4 vibrio chol
19	49	36.0	1047	3	Q8TFZ1	Q8tfz1 aspergillus
20	48.5	35.7	311	16	Q9PIC9	Q9pic9 campylobact
21	48.5	35.7	688	5	O96497	O96497 trypanosoma
22	48	35.3	225	17	Q96250	Q96250 sulfolobus
23	48	35.3	235	17	Q976P7	Q976p7 sulfolobus
24	48	35.3	330	16	Q9K7G4	Q9k7g4 bacillus ha
25	48	35.3	609	10	Q9LV35	Q9lv35 arabidopsis
26	48	35.3	639	11	Q8BIN9	Q8bin9 mus musculu
27	48	35.3	1034	3	Q9Y7E9	Q9y7e9 candida gla
28	48	35.3	1659	4	Q8IWQ9	Q8iwq9 homo sapien
29	47.5	34.9	444	16	Q8XXK5	Q8xxk5 clostridium
30	47	34.6	161	2	Q8GJ63	Q8gj63 bartonella
31	47	34.6	319	10	O81460	O81460 arabidopsis
32	47	34.6	352	17	Q980E9	Q980e9 sulfolobus
33	47	34.6	399	16	Q8PIN0	Q8pin0 streptococc
34	47	34.6	476	10	Q8RVV4	Q8rvv4 lycopersico
35	47	34.6	581	11	Q922Q3	Q922q3 mus musculu
36	47	34.6	662	16	Q8Z7U3	Q8z7u3 salmonella
37	47	34.6	732	16	Q8EAI4	Q8eai4 shewanella
38	47	34.6	1072	3	Q8WZX5	Q8wzx5 neurospora
39	47	34.6	1261	5	O8MMT3	Q8mmt3 dictyosteli
40	47	34.6	2388	5	Q8MYG5	Q8myg5 dictyosteli
41	47	34.6	2473	12	O71209	O71209 grapevine 1
42	46.5	34.2	541	16	Q8EFD3	Q8efd3 shewanella
43	46.5	34.2	614	5	Q22551	Q22551 caenorhabdi
44	46.5	34.2	771	5	Q19380	Q19380 caenorhabdi
45	46	33.8	320	5	Q8IJX6	Q8ijx6 plasmodium

ALIGNMENTS

RESULT 1
Q9U1Q4
ID Q9U1Q4 PRELIMINARY; PRT; 1050 AA.
AC Q9U1Q4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Y87G2A.5 protein.
GN Y87G2A.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL110500; CAB60428.1; -.
DR HSSP; P96142; IGAX.
DR WormPep; Y87G2A.5; CE24685.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_I.
DR InterPro; IPR002303; tRNA-synt_val.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00986; TRNASYNTHVAL.
DR TIGRFAMS; TIGR00422; vals; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
SQ SEQUENCE 1050 AA; 118920 MW; F33DB53587EAC057 CRC64;

Query Match 40.4%; Score 55; DB 5; Length 1050;
Best Local Similarity 44.4%; Pred. No. 22;
Matches 12; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
SQ SEQUENCE 1049 AA; 118253 MW; 13A513ABF69E8EEB CRC64;

Query Match 37.5%; Score 51; DB 5; Length 1049;
Best Local Similarity 40.7%; Pred. No. 89;
Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQGVETIVSFH 28
|| :|| | ||| |::| |
Db 471 YVSCSDMAASATEAVRSGELKIIPEHH 497

RESULT 4
Q960E6 PRELIMINARY; PRT; 1049 AA.
AC Q960E6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SD04748p.
GN AATS-VAL OR CG4062.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY052099; AAK93523.1; -
DR FlyBase; FBgn0027079; Aats-val.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002303; tRNA-synt_val.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00986; TRNASYNTHVAL.
DR TIGRFAMS; TIGR00422; vals; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
SQ SEQUENCE 1049 AA; 118331 MW; 56F322C7414EEAC4 CRC64;

Query Match 37.5%; Score 51; DB 5; Length 1049;
Best Local Similarity 40.7%; Pred. No. 89;
Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQGVETIVSFH 28
|| :|| | ||| |::| |
Db 471 YVSCSDMAASATEAVRSGELKIIPEHH 497

RESULT 5
O82521 PRELIMINARY; PRT; 459 AA.
AC O82521;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative aminotransferase.
OS Capsicum chinense.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=80379;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=habanero;
RA Aluru M., Curry J., O'Connell M.;
RT "Nucleotide Sequence of a Probable Aminotransferase Gene (Accession
No. AF085149) from Habanero Chile. (PGR98-182).";

RL Plant Physiol. 118:1102-1102(1998).
DR EMBL; AF085149; AAC78480.1; -.
DR HSSP; P04181; 2OAT.
DR InterPro; IPR005814; Aminotrans_3.
DR Pfam; PF00202; aminotran_3; 1.
DR PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
KW Aminotransferase; Transferase.
SQ SEQUENCE 459 AA; 50729 MW; 02ABB4D728B524E4 CRC64;

Query Match 36.8%; Score 50; DB 10; Length 459;
Best Local Similarity 41.7%; Pred. No. 51;
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQGVETIVSF 27
||| :|| | || |::| |
Db 196 FSTRLANLESLILKEGPETVAAF 219

RESULT 6
Q9H6R2 PRELIMINARY; PRT; 642 AA.
AC Q9H6R2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein FLJ21965.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025618; BAB15191.1; -
DR HSSP; P96142; 1GAX.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR002303; tRNA-synt_val.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00986; TRNASYNTHVAL.
KW Hypothetical protein.
SQ SEQUENCE 642 AA; 71578 MW; C9E37EED742B7F1 CRC64;

Query Match 36.8%; Score 50; DB 4; Length 642;
Best Local Similarity 44.4%; Pred. No. 73;
Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQGVETIVSFH 28
:| || :||:| | :| |||
Db 60 FVRCQEMGARAKAVESGALELSPSFH 86

RESULT 7
Q96GN2 PRELIMINARY; PRT; 642 AA.
AC Q96GN2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009355; AAH09355.1; -.

```
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR002303; tRNA-synt_val.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00986; TRNASYNTHVAL.
KW Hypothetical protein.
SQ SEQUENCE 642 AA; 71650 MW; 1F00CBB73742B579 CRC64;

Query Match 36.8%; Score 50; DB 4; Length 642;
Best Local Similarity 44.4%; Pred. No. 73;
Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQVETIVSFH 28
:| || :||:| | :| |||
Db 60 FVRCQEMGARAARAKAVESGALELSPSFH 86

RESULT 8
Q96Q02 PRELIMINARY; PRT; 1098 AA.
AC Q96Q02;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA1885 (Fragment).
GN KIAA1885.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins."
RL DNA Res. 8:179-187(2001).
DR EMBL; AB067472; BAB67778.1; -
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_I.
DR InterPro; IPR002303; tRNA-synt_val.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00986; TRNASYNTHVAL.
DR TIGRFAMS; TIGR00422; vals; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1098 AA; 122470 MW; E01DCA8C8E42BC4D CRC64;

Query Match 36.8%; Score 50; DB 4; Length 1098;
Best Local Similarity 44.4%; Pred. No. 1.3e+02;
Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 2 YVFSTEMANKAAEAVLKQVETIVSFH 28
:| || :||:| | :| |||
Db 516 FVRCQEMGARAARAKAVESGALELSPSFH 542

RESULT 9
Q8GGI9 PRELIMINARY; PRT; 350 AA.
AC Q8GGI9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Abc1.
GN Abc1.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=CCM3626;
RA Bringel F., Hubert J.-C.;
RT "Investigation of arginine requirement in Lactobacillus isolated from
RT different environments revealed point mutations, insertions and
RT deletions in arginine biosynthetic genes."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF514870; AA015990.1; -
SQ SEQUENCE 350 AA; 38178 MW; 64D1986ED73C8AAA CRC64;

Query Match 36.4%; Score 49.5; DB 2; Length 350;
Best Local Similarity 54.5%; Pred. No. 45;
Matches 12; Conservative 4; Mismatches 3; Indels 3; Gaps 1;

QY 3 VFSTEMANKAAEAVLKQVETI 24
|:|::| ||| |||:|
Db 176 VYSTDLLAKAAE--KGQVDAI 194

RESULT 10
Q9LFW7 PRELIMINARY; PRT; 141 AA.
AC Q9LFW7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T7N9.31.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T7N9 from chromosome
RT I."
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howng B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howng B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
```


RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.,
RL Submitted (SEP-2000) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AC000348; AAF79866.1; --
SQ SEQUENCE 141 AA; 16649 MW; 6FEC48B07326B313 CRC64;

Query Match 36.0%; Score 49; DB 10; Length 141;
Best Local Similarity 41.7%; Pred. No. 19;
Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQGVETIVSF 27
: | | | | : | | : | | : | |
Db 112 WFSRSTNKAADRLAKGELENNVTF 135
: | | | | : | | : | | : | |

RESULT 11
Q976C9
ID Q976C9 PRELIMINARY; PRT; 143 AA.
AC Q976C9;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Hypothetical protein St0252.
GN ST0252.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000982; BAB65218.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 143 AA; 16822 MW; D1AFCC5157298CA2 CRC64;

Query Match 36.0%; Score 49; DB 17; Length 143;
Best Local Similarity 43.3%; Pred. No. 20;
Matches 13; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

QY 1 VYVFSTEMAN--KAAEAVLKQGVETIVSFH 28
| | | | | | | | | | | | : | |
Db 26 VLVVLTNMKNVEKEAKVLKTRIDKVVYIH 55
: | | | | | | | | | | | | : | |

RESULT 12
Q8G3S3
ID Q8G3S3 PRELIMINARY; PRT; 187 AA.
AC Q8G3S3;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypoxanthine-guanine phosphoribosyltransferase.
GN HPRT OR BL1681.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,

RA Pridmore R.D., Arigoni F.,
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AE014802; AAN25468.1; --
KW Glycosyltransferase; Transferrase; Complete proteome.
SQ SEQUENCE 187 AA; 20617 MW; E697C3C127277DB1 CRC64;

Query Match 36.0%; Score 49; DB 16; Length 187;
Best Local Similarity 45.2%; Pred. No. 26;
Matches 14; Conservative 3; Mismatches 4; Indels 10; Gaps 1;

QY 7 EMANKAAE-----AVLKQGVETIVSF 27
| | | | | : | | | | | : | |
Db 26 EMAALASEDYRDKNPDLLVAVLKGAVENTLVAF 56
: | | | | | : | | | | | : | |

RESULT 13
Q9FZL1
ID Q9FZL1 PRELIMINARY; PRT; 210 AA.
AC Q9FZL1;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE F17L21.2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Johnson-Hopson C., Brooks S., Buehler E., Chao Q., Khan S., Kim C.,
RA Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F17L21 from chromosome
RT I.";
RL Submitted (APR-1998) to the EMBL/GenBank/DBDJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBDJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBDJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBDJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBDJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AC004557; AAF99722.1; --
DR InterPro; IPR003822; PAH.
DR Pfam; PF02671; PAH; 1.
SQ SEQUENCE 210 AA; 24697 MW; FCD8130CD75700A0 CRC64;


```

Query Match          36.0%; Score 49; DB 10; Length 210;
Best Local Similarity 41.7%; Pred. No. 30;
Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

```

Qy	4	FSTEMANKAAEAVLKQVETIVSF	27
	:		:
Dp	181	WSEFSTNKAADRLAKGELENNVTF	204

RESULT 14

Q96YA9
ID Q96YA9 PRELIMINARY; PRT; 233 AA.

Q96YA9; TREMBLrel. 19, Created)
01-DEC-2001 (TREMBLrel. 19, Last sequence update)
01-DEC-2001 (TREMBLrel. 19, Last annotation update)
Hypothetical protein ST2259.

OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.

NCBI TaxID=111955;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JCM 10545 / 7;

RC [SIRKIN-UCM 10543](#)
RX [PubMed=11572479](#):

RA Kwararabasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;

RT "Complete genome sequence of an aerobic thermoacidophilic
RT *Crenarchaeon*, *Sulfolobus tokodaii* strain 7.";
RT *Crenarchaeon*, *Sulfolobus tokodaii* strain 7.";

RI *Ctenactis*, *Stictopora* *lokooe*
BL DNA Res. 8:123-140(2001).

DR EMBL: AP000989: BAB67368.1; -.
DNA RES: 0.125 140(2001).

KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 233 AA; 27253 MW; 8B9E6FDB60EEEE36D CRC64;

QY 1 VYVFSTEMANKAAEAVLKQVETIVSFH 28
| | | : | | | | : : | |
Db 120 VLVLTKNVEKEAEKYLKTRIDKVVYIH 14

RESULT 15

Q96X55

ID Q96X55 PRELIMINARY; PRT; 237 AA.

AC Q96X55;

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLREL. 19, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Hypothetical protein ST1165.

GN ST1165 OR ST0857.

OS Sulfolobus tokodaii.

OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

OC Sulfolobus.

OX NCBI TaxID=111955;

RN [11]

RP SEQUENCE FROM N. A.

RE SEQUENCE FROM N.A.
RC STRATN² JCM 10545 / 7.

RX STRAIN=JCM 10343
PubMed=11572479.

RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Fumed-11372479;
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;

RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain 7.";

[illegible]

Search completed: October 15, 2003, 10:35:44
Job time : 19.2335 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 119.268 Seconds
(without alignments)
258.182 Million cell updates/sec

Title: US-09-915-543-15_COPY_199_392

Perfect score: 1028

Sequence: 1 TIVSFHIQNISNNKTERSTA.....LFPDEKEFTGAQSGGPOQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	A_Geneseq_19Jun03:*
1:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1028	100.0	1426	23 AAB71229	Human legless homo
2	1028	100.0	1435	22 ABB11808	Human BCL9 homolog
3	277	26.9	140	22 AAO05855	Human polypeptide
4	232.5	22.6	320	23 AAU78461	Mouse beta-catenin
5	232.5	22.6	1494	23 AAU78460	Mouse beta-catenin
6	167.5	16.3	707	22 AAM39141	Human polypeptide
7	162.5	15.8	574	22 AAM52322	WASP homolog prote
8	162.5	15.8	574	22 AAG67370	Amino acid sequenc
9	159	15.5	406	22 ABG27250	Novel human diagno

10	158.5	15.4	256	22 ABU52945	Human mammary carc
11	158.5	15.4	270	22 ABU52939	Human mammary carc
12	158.5	15.4	446	22 ABB70063	Drosophila melanog
13	158	15.4	783	19 AAW37151	Mouse neural Mena+
14	158	15.4	787	19 AAW37152	Mouse neural Mena+
15	158	15.4	802	19 AAW37153	Mouse neural Mena+
16	158	15.4	802	22 AAU09139	Mammalian enabled
17	157	15.3	598	22 ABG14000	Novel human diagno
18	155	15.1	253	22 ABU52937	Human mammary carc
19	155	15.1	572	18 AAW31855	Mycobacterium tube
20	155	15.1	763	18 AAW31852	Mycobacterium tube
21	154.5	15.0	731	22 AAB74209	Protein encoded by
22	154.5	15.0	731	23 ABB93202	Herbicidally activ
23	153.5	14.9	177	22 ABU52946	Human mammary carc
24	151.5	14.7	441	22 ABB11413	Human extensin hom
25	151	14.7	693	23 ABP69529	Human polypeptide
26	150.5	14.6	533	16 AAR72483	Human H-2RIIBp. H
27	150.5	14.6	533	23 ABB79960	Retinoid C recepto
28	150.5	14.6	533	23 AAO19271	Human retinoid X r
29	150.5	14.6	533	23 AAO19282	Human retinoid X r
30	150.5	14.6	533	23 AAO19283	Human retinoid X r
31	150.5	14.6	533	23 AAO19284	Human retinoid X r
32	150.5	14.6	533	23 AAO19285	Human retinoid X r
33	150.5	14.6	533	23 AAO19286	Human retinoid X r
34	150.5	14.6	533	23 AAO19287	Human retinoid X r
35	150.5	14.6	533	23 AAO19288	Human retinoid X r
36	150.5	14.6	533	23 AAO19289	Human retinoid X r
37	150.5	14.6	533	23 AAO19291	Human retinoid X r
38	150.5	14.6	533	23 AAO19292	Human retinoid X r
39	150.5	14.6	533	23 AAO19293	Human retinoid X r
40	150.5	14.6	533	23 AAO19294	Human retinoid X r
41	150.5	14.6	533	23 AAO19295	Human retinoid X r
42	150.5	14.6	533	23 AAO19296	Human retinoid X r
43	150.5	14.6	533	23 AAO19297	Human retinoid X r
44	150.5	14.6	533	23 AAO19298	Human retinoid X r
45	150.5	14.6	533	23 AAO19299	Human retinoid X r

ALIGNMENTS

RESULT 1

AAB71229

ID AAB71229 standard; Protein; 1426 AA.

XX

AC AAB71229;

XX

DT 18-NOV-2002 (first entry)

XX

DE Human legless homologue lgs/bcl9 protein.

XX

KW Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;

KW tissue proliferation; tumour; cytostatic; cellular disorder; colon;

KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;

KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.

XX

OS Homo sapiens.

XX

PN US2002086986-A1.

XX

PD 04-JUL-2002.

XX

PF 27-JUL-2001; 2001US-0915543.

XX

PR 28-JUL-2000; 2000US-221502P.

XX

PA (BASL/) BASLER K.

PA (BRUN/) BRUNNER E.

PA (FROE/) FROESCH B.

PA (KRAM/) KRAMPS T.

PA (PETE/) PETER O.

XX

PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;

QY 121 QGSNSSADPKAPPPPPVSSGEPPTLGENPDGLSQEQLHRSLSQTLRDQIRMLFPDEK 180
Db 359 QGSNSSADPKAPPPPPVSSGEPPTLGENPDGLSQEQLHRSLSQTLRDQIRMLFPDEK 418
QY 181 EFTGAQSGGPQQNP 194
Db 419 EFTGAQSGGPQQNP 432
RESULT 3
AAO05855
ID AAO05855 standard; Protein; 140 AA.
XX AC AAO05855;
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 19747.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US04927.
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX N-PSDB; AAI85786.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX PS Claim 20; SEQ ID NO 19747; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 140 AA;

Query Match 26.9%; Score 277; DB 22; Length 140;
Best Local Similarity 91.4%; Pred. No. 1.7e-12;
Matches 53; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 TIVSFHQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQIP 58
Db 81 TMVSFHIQNISSSKTERSTAPLNTQISALRNDPKPLPQQPPAPASQDNYSQNTRLQP 138
XX

RESULT 4
AAU78461
ID AAU78461 standard; Protein; 320 AA.
XX AC AAU78461;
XX DT 02-JUL-2002 (first entry)
XX DE Mouse beta-catenin nuclear localised protein #2.
XX KW Mouse; beta-catenin nuclear localised protein; cancer;
KW gene therapy; EST; expressed sequence tag.
XX OS Mus musculus.
XX PN WO200224738-A1.
XX PD 28-MAR-2002.
XX PF 19-SEP-2001; 2001WO-JP08140.
XX PR 22-SEP-2000; 2000JP-0287876.
XX PA (KYOW) KYOWA HAKKO KOGYO KK.
XX PI Akiyama T, Adachi S;
XX WPI; 2002-330014/36.
XX N-PSDB; ABK47632.
XX New beta-catenin nuclear localised protein for diagnosis and treatment
PT of diseases associated with nuclear localisation of beta-catenin e.g.
PT cancer -
XX PS Claim 2; Page 91-92; 113pp; Japanese.
XX CC The invention relates to a beta-catenin nuclear localised protein
CC and DNA encoding the protein. The protein and encoding DNA are
CC applicable in diagnosis and treatment of diseases associated with
CC nuclear localisation of beta-catenin e.g. cancer, including gene
CC therapy. The present sequence represents the amino acid sequence of
CC mouse beta-catenin nuclear localised protein #2.
XX SQ Sequence 320 AA;

Query Match 22.6%; Score 232.5; DB 23; Length 320;
Best Local Similarity 35.1%; Pred. No. 5.7e-09;
Matches 71; Conservative 23; Mismatches 75; Indels 33; Gaps 9;

QY 1 TIVSFHQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQPTP 60
Db 16 SILAYHQNVPRAKLDQA-----PKVPTPEPLPLN---TPSAGTPQSQPP 58
QY 61 PIPAPAPKPAAPRPLDRESP--GVENKLIP-SVGSPASSTPLPPDGTGPNSTPNRAVT 117
Db 59 PLPPPPPPAPGSAPPALPPEGPPEDTSQDLAPNSVG--AASGGGTGGTHPNTPTAATANN 116
QY 118 PVSQGSNSSADPKA--PPPPPVSSGEPPTLGENPDGLSQEQLHRSLSQTLRDQIRML 175
Db 117 PLPPGGDPGSAPGSALLGEATPTGNGQRNLVGS--EGLSKQLEHRSLSQTLRDIERLL 174
QY 176 F--PDEKEFT-----GAQSGGP 190
Db 175 LRSGETEFLKGPPEGAGEGP 196

RESULT 5
AAU78460
ID AAU78460 standard; Protein; 1494 AA.
XX AC AAU78460;
XX

XX Human; gene therapy; vaccine; disease treatment; detection.

XX Homo sapiens.

XX WO200112659-A2.

XX 22-FEB-2001.

XX 18-AUG-2000; 2000WO-IB01496.

XX 18-AUG-1999; 99US-0149499.

XX 28-SEP-1999; 99US-0156503.

XX (GEHU-) GERMAN HUMAN GENOME PROJECT.

XX Wiemann S;

XX WPI; 2001-327840/34.

XX Nucleic acids having the sequences of clones isolated from libraries of different human tissues, useful in recombinant DNA methodologies -

XX Example III; Page 550; 1095pp; English.

XX This invention describes novel polynucleotides and polypeptides isolated from human cDNA libraries which can be used for gene therapy or in vaccines. The polynucleotides of the invention and antibodies encoded by them may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate polypeptide expression. The products of the invention may also be used to identify modulators of expression and activity and to down regulate expression and activity. The antibodies of the invention may also be used as diagnostic agents for detecting the presence of polypeptides in samples. This sequence represents a homologue of a polypeptide described in the disclosure of the invention.

XX Sequence 270 AA;

Query Match 15.4%; Score 158.5; DB 22; Length 270;

Best Local Similarity 32.0%; Pred. No. 0.00087;

Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;

QY 14 KTERSTAPLNTQISALRNDPKPLP--QQPPAPANQDNSSQNTRLQP---TPPIAPAPK 68

Db 79 KSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKSSPPPPAPMSSPPPPPEVKSSPPPPAPVSS 138

QY 69 P-----AAPRRPLDRSPGVENKLIPSVGSPASSTPLP---PDGTGNSTPNRAVTPV 119

Db 139 PPPPVKSPPPAPVSSPPPVKS---PPPPAPVSSPPPPVKSPPPAPISSPPPPVKSP 195

QY 120 SQGSNSSADP-KAPPPP-PVSSGEPTLGENP 150

Db 196 PPAPVSSPPPPVKSPPPAPVSSPPPIKSP 228

RESULT 12

ABB70063

ID ABB70063 standard; Protein; 446 AA.

XX ABB70063;

XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster polypeptide SEQ ID NO 36981.

DE Drosophila; developmental biology; cell signalling; insecticide;

XX Drosophila melanogaster.

OS Drosophila melanogaster.

XX WO200171042-A2.

PN 27-SEP-2001.

PD

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR N-PSDB; ABL14166.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX Disclosure; SEQ ID NO 36981; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 446 AA;

Query Match 15.4%; Score 158.5; DB 22; Length 446;

Best Local Similarity 29.6%; Pred. No. 0.0014;

Matches 42; Conservative 11; Mismatches 64; Indels 25; Gaps 3;

QY 33 PKPLPQQPPAPANQDNSSQNTRLQTPPIPAPA--PKPAAPRPLDRSPGVENKLIPS 90

Db 97 PQTRPPPPPPPP-----QPTPPAPRPSYGGPPQTQPPRPPPTPSAPAPPPPS 144

QY 91 VG-----SPASSTPLPDGTGNSTPNRAVTPVSQGSNSSADPKAPPPPPVS 139

Db 145 YGPPQTTPPPRPPQTTPSAPAPPPSYGPPQTTPPPRPPPTPSAPAPSYGPPQPPAPQ 204

QY 140 SGEPTLGENPDGLSQEQLEHR 161

Db 205 PPSPPSQPGPEYLPDPDPKPR 226

RESULT 13

AAW37151

ID AAW37151 standard; Protein; 783 AA.

XX AAW37151;

XX 06-JUL-1998 (first entry)

DT Mouse neural Mena+ protein.

DE Neural Mena+ protein; mammalian Ena; Enabled protein; Evi protein; cytoskeleton; cell morphology; cell adhesion; cell differentiation; cell growth; cell motility; mouse.

XX Mus musculus.

XX Key Location/Qualifiers

FT Misc-difference 378

FT /note= "encoded by GGN"

XX WO9801755-A1.

XX 15-JAN-1998.

PD

XX PF 03-JUL-1997; 97WO-US11669.
XX PR 05-JUL-1996; 96US-0675815.
XX PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PI Gertler FB, Niebuhr K, Soriano P, Wehland J;
XX WPI; 1998-101197/09.
XX DR N-PSDB; AAV02998.
XX PT Detection of modulators of Mena and Ena-VASP-like genes and proteins
XX PT - used in control of cytoskeletal dynamic events in normal and
XX PT abnormal cell morphology, adhesion, motility, growth and
XX PT differentiation
XX PS Example 4; Page 58-60; 77pp; English.
XX CC This protein comprises novel murine neural Mena+. Its amino acid
XX CC sequence was deduced from a cDNA clone (see AAV02998) obtained from
XX CC a mouse brain cDNA library. Neural Mena+ contains an exon that
XX CC introduces 244 amino acids between amino acids 238 and 239 of
XX CC mammalian Ena (Mena, see AAW37148). Two other isoforms, neural
XX CC Mena++ (see AAW37152) and neural Mena+++ (see AAW37153), are also
XX CC disclosed. Unlike Mena, neural Mena isoforms exhibit neural
XX CC tissue-specific distribution. Based on the disclosed Mena and Evl
XX CC genes (see also AAV02996-97) and proteins (see also AAW37148-49), a
XX CC variety of methods and compositions are provided for screening,
XX CC isolating and characterising endogenous and exogenous factors,
XX CC drugs and therapeutic agents useful to evaluate and/or control
XX CC cytoskeletal dynamic events involved in normal and abnormal cell
XX CC morphology, adhesion, motility, growth and/or differentiation. A
XX CC method of detecting a modulator of Mena activity/expression is
XX CC claimed.
XX SQ Sequence 783 AA;
Query Match 15.4%; Score 158; DB 19; Length 783;
Best Local Similarity 26.3%; Pred. No. 0.0028;
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;
QY 13 NKTERSTAPLNT-----QISALRNDP----- 33
Db 329 NKNSRPSPVNTPSQPPAAKSCAWPTSNFSPLPSPPIMISSPPGKATGPRVLPVCVS 388
QY 34 KPLPQQPPAPANQDN-----SSQNTRLQPTPIPAAPKPAAPRPL----- 76
Db 389 SPVPQMPSPPTAPNGSLDSVTYPVSPPTSGPAAPPPPPPPPPPPPLPPLPLAS 448
QY 77 -----DRESPGVENKLI-----PSVGSPPASS-TPLPPD-GTGNSTPNRAV- 116
Db 449 LSHCGSQASPPPGTPLASTPSSKPSVLPSPSAGAPASAEPLNPELGDSSASEPGLQAAS 508
QY 117 ----TPVSQGSNSSADPKAPPPPPVSSCE-----PPTLGENP 150
Db 509 QPAESPTPQG--LVLGPPAPPPPPPLPSGPPAYASALPPPPGPPP 550
RESULT 14
AAW37152
ID AAW37152 standard; Protein; 787 AA.
XX AC AAW37152;
XX XX
XX DT 06-JUL-1998 (first entry)
XX DE Mouse neural Mena++ protein.
XX XX
KW Neural Mena++ protein; mammalian Ena; Enabled protein; Evl protein;
KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
KW cell growth; cell motility; mouse.

XX OS Mus musculus.
XX PN WO9801755-A1.
XX PD 15-JAN-1998.
XX PF 03-JUL-1997; 97WO-US11669.
XX PR 05-JUL-1996; 96US-0675815.
XX PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PI Gertler FB, Niebuhr K, Soriano P, Wehland J;
XX WPI; 1998-101197/09.
XX PT Detection of modulators of Mena and Ena-VASP-like genes and proteins
XX PT - used in control of cytoskeletal dynamic events in normal and
XX PT abnormal cell morphology, adhesion, motility, growth and
XX PT differentiation
XX PS Example 4; Page 60-63; 77pp; English.
XX CC This protein comprises novel murine neural Mena+. Its amino acid
XX CC sequence was deduced from a cDNA clone obtained from a mouse brain
XX CC cDNA library. Two other isoforms, neural Mena+ (see AAW37151) and
XX CC neural Mena+++ (see AAW37153), are also disclosed. Unlike mammalian
XX CC Ena (Mena, see AAW37148), neural Mena isoforms exhibit neural
XX CC tissue-specific distribution. Based on the disclosed Mena and Evl
XX CC genes (see also AAV02996-98) and proteins (see also AAW37148-49), a
XX CC variety of methods and compositions are provided for screening,
XX CC isolating and characterising endogenous and exogenous factors,
XX CC drugs and therapeutic agents useful to evaluate and/or control
XX CC cytoskeletal dynamic events involved in normal and abnormal cell
XX CC morphology, adhesion, motility, growth and/or differentiation. A
XX CC method of detecting a modulator of Mena activity/expression is
XX CC claimed.
XX SQ Sequence 787 AA;
Query Match 15.4%; Score 158; DB 19; Length 787;
Best Local Similarity 26.3%; Pred. No. 0.0028;
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;
QY 13 NKTERSTAPLNT-----QISALRNDP----- 33
Db 333 NKNSRPSPVNTPSQPPAAKSCAWPTSNFSPLPSPPIMISSPPGKATGPRVLPVCVS 392
QY 34 KPLPQQPPAPANQDN-----SSQNTRLQPTPIPAAPKPAAPRPL----- 76
Db 393 SPVPQMPSPPTAPNGSLDSVTYPVSPPTSGPAAPPPPPPPPPPPPLPPLPLAS 452
QY 77 -----DRESPGVENKLI-----PSVGSPPASS-TPLPPD-GTGNSTPNRAV- 116
Db 453 LSHCGSQASPPPGTPLASTPSSKPSVLPSPSAGAPASAEPLNPELGDSSASEPGLQAAS 512
QY 117 ----TPVSQGSNSSADPKAPPPPPVSSGE-----PPTLGENP 150
Db 513 QPAESPTPQG--LVLGPPAPPPPPPLPSGPPAYASALPPPPGPPP 554
RESULT 15
AAW37153
ID AAW37153 standard; Protein; 802 AA.
XX AC AAW37153;
XX XX
XX DT 06-JUL-1998 (first entry)
XX XX
DE Mouse neural Mena+++ protein.
XX XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 39.2529 Seconds
(without alignments)
209.113 Million cell updates/sec

Title: US-09-915-543-15_COPY_199_392
Perfect score: 1028
Sequence: 1 TIVSFHIQNISNKNTERSTA.....LPDEKEFTGAQSGGPOQNP 194

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	161.5	15.7	816	4	US-09-266-225D-12	Sequence 12, Appl
2	150.5	14.6	533	1	US-07-952-800-2	Sequence 2, Appl
3	149.5	14.5	525	3	US-08-764-870-7	Sequence 7, Appl
4	149.5	14.5	525	3	US-08-980-115-7	Sequence 7, Appl
5	143.5	14.0	446	1	US-07-952-800-4	Sequence 4, Appl
6	140.5	13.7	8991	4	US-08-714-741-32	Sequence 32, Appl
7	138.5	13.5	297	2	US-08-580-545B-6	Sequence 6, Appl
8	138.5	13.5	297	3	US-09-262-653A-6	Sequence 6, Appl
9	137	13.3	456	2	US-08-819-013-1	Sequence 1, Appl
10	136.5	13.3	142	4	US-09-252-991A-24873	Sequence 24873, A
11	136.5	13.3	1248	2	US-09-080-897-2	Sequence 2, Appl
12	136.5	13.3	1248	3	US-09-323-735-2	Sequence 2, Appl
13	135	13.1	332	3	US-08-818-112-53	Sequence 53, Appl
14	135	13.1	332	4	US-08-818-111-53	Sequence 53, Appl
15	135	13.1	332	4	US-09-056-556-53	Sequence 53, Appl
16	135	13.1	332	4	US-09-072-596-53	Sequence 53, Appl
17	135	13.1	334	6	5202236-3	Patent No. 5202236
18	133.5	13.0	941	4	US-07-757-022B-14	Sequence 14, Appl
19	133.5	13.0	1022	4	US-07-757-022B-84	Sequence 84, Appl
20	133.5	13.0	1038	4	US-07-757-022B-74	Sequence 74, Appl
21	133.5	13.0	1049	4	US-07-757-022B-58	Sequence 58, Appl
22	133.5	13.0	1140	4	US-07-757-022B-104	Sequence 104, App
23	133.5	13.0	1270	4	US-07-757-022B-44	Sequence 44, Appl
24	133.5	13.0	1311	4	US-07-757-022B-42	Sequence 42, Appl
25	133.5	13.0	1313	4	US-07-757-022B-142	Sequence 142, App
26	133.5	13.0	1314	4	US-07-757-022B-50	Sequence 50, Appl
27	133.5	13.0	1320	4	US-07-757-022B-46	Sequence 46, Appl

28	133.5	13.0	1320	4	US-07-757-022B-60	Sequence 60, Appl
29	133.5	13.0	1354	4	US-07-757-022B-48	Sequence 48, Appl
30	133.5	13.0	1361	4	US-07-757-022B-40	Sequence 40, Appl
31	133.5	13.0	1363	4	US-07-757-022B-52	Sequence 52, Appl
32	133.5	13.0	1404	4	US-07-757-022B-2	Sequence 2, Appl
33	133.5	13.0	1404	4	US-07-757-022B-62	Sequence 62, Appl
34	133	12.9	1315	3	US-08-899-595-3	Sequence 3, Appl
35	132	12.8	214	1	US-08-217-327-4	Sequence 4, Appl
36	132	12.8	331	6	5202236-37	Patent No. 5202236
37	132	12.8	1231	4	US-08-714-741-41	Sequence 41, Appl
38	131	12.7	2442	4	US-09-514-247A-10	Sequence 10, Appl
39	130.5	12.7	325	1	US-08-382-184-2	Sequence 2, Appl
40	130.5	12.7	325	2	US-08-641-356-2	Sequence 2, Appl
41	130.5	12.7	325	3	US-09-132-528-2	Sequence 2, Appl
42	130.5	12.7	325	3	US-09-132-528-3	Sequence 2, Appl
43	130.5	12.7	325	3	US-08-875-494-2	Sequence 2, Appl
44	130.5	12.7	325	4	US-09-599-366-2	Sequence 2, Appl
45	130.5	12.7	325	4	US-09-599-366-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-266-225D-12
; Sequence 12, Application US/09266225D
; Patent No. 6573364
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishan
; APPLICANT: Kingsmore, Stephen
; APPLICANT: Tchernev, Velizar
; TITLE OF INVENTION: Isolation and Characterization of Hermansky-Pudlak
; TITLE OF INVENTION: Syndrome (HPS) Protein Complexes and HPS Protein-
; TITLE OF INVENTION: Interacting Proteins
; FILE REFERENCE: 15966-523
; CURRENT APPLICATION NUMBER: US/09/266,225D
; CURRENT FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 816
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-266-225D-12

Query Match	15.7%;	Score	161.5;	DB	4;	Length	816;
Best Local Similarity	32.1%;	Pred. No.	1.6e-05;				
Matches	54;	Conservative	7;	Mismatches	64;	Indels	43;
						Gaps	7;
QY	17	RSTAPLNTQISALRNDPKPLP-----	-----QOPPAPANQDNSSQNRLQTPPIP	63			
Db	577	RPAAPALTSVPAPAPAPTPTTPVQTSPPPGPVAGPTGPQPQSAGSTSGVPQPACPPP	636				
QY	64	APAPKPAAPRPLDRESPG-----	-----VENKLIPSVGSPASSTP-----LPPDGT	105			
Db	637	GPAPHTGPPGPIPVPAAPPQIATSTLLAAQSLVPPPGGLPGSSTGVLPFPGLPPDA	696				
QY	106	G--PNS-----TPNNRAVTPVSGSNSSSADPKAPPPPPVSSGEPPTLG	147				
Db	697	GGAPQSSMSESPDVNLVT--QQLSKSQVEDPL----	PPVFSGTPKGSG 738				

RESULT 2
US-07-952-800-2
; Sequence 2, Application US/07952800
; Patent No. 5403925
; GENERAL INFORMATION:
; APPLICANT: OZATO, KEIKO
; TITLE OF INVENTION: A NEW MEMBER OF THE NUCLEAR HORMONE
; TITLE OF INVENTION: RECEPTOR SUPERFAMILY AND A CDNA CLONE THEREOF
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND

Db 6056 --APAPAPAPKPKPAEKPAKPAPAPKPKPTPTKTGWKQE 6090

RESULT 7

US-08-580-545B-6
; Sequence 6, Application US/08580545B
; Patent No. 5932713
; GENERAL INFORMATION:
; APPLICANT: Yoshihisa, Kasukabe
; APPLICANT: Koichi, Fujisawa
; APPLICANT: Susumu, Nishiguchi
; APPLICANT: Yoshihiko, Maekawa
; APPLICANT: Randy, Allen
; TITLE OF INVENTION: COTTON FIBER TISSUE-SPECIFIC GENES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 601 Thirteenth Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/580,545B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bretschneider, Barry E.
; REGISTRATION NUMBER: 28,055
; REFERENCE/DOCKET NUMBER: 04473/068001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/783-5070
; TELEFAX: 202/783-2331
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-580-545B-6

Query Match 13.5%; Score 138.5; DB 2; Length 297;
Best Local Similarity 31.2%; Pred. No. 0.00038;
Matches 43; Conservative 9; Mismatches 59; Indels 27; Gaps 6;

QY 15 TERSTAPLNTQISALRNDPKPLPQQPPAPAPANQDNSSQNTRLQPTPPI--PAPAPKPAAP 72
Db 78 TKAPTPPYKPPPTKA---PTP-PYKPPAPAPPTK-----APTPPYKPPAPAPPTKAP 124
QY 73 PRPLDRESPGVENKLIPSVGSPASSTPLPPDGTGPNSTPNNRVTPVSGSNSSSADPKA 132
Db 125 TPTYKPPAP-----APPTKAPTPP-FKPPAPAPPTKAPTPPYKPPPTPAPAPPVK 172
QY 133 PPPPPVSSGEPPTLGENP 150
Db 173 APTPPYMPPTPTKAPT 190

RESULT 8

US-09-262-653A-6
; Sequence 6, Application US/09262653A
; Patent No. 6166294
; GENERAL INFORMATION:
; APPLICANT: Yoshihisa, Kasukabe
; APPLICANT: Koichi, Fujisawa
; APPLICANT: Susumu, Nishiguchi
; APPLICANT: Yoshihiko, Maekawa
; APPLICANT: Randy, Allen

; TITLE OF INVENTION: COTTON FIBER TISSUE-SPECIFIC GENES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 601 Thirteenth Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/262,653A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bretschneider, Barry E.
; REGISTRATION NUMBER: 28,055
; REFERENCE/DOCKET NUMBER: 04473/068001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/783-5070
; TELEFAX: 202/783-2331
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-262-653A-6

Query Match 13.5%; Score 138.5; DB 3; Length 297;
Best Local Similarity 31.2%; Pred. No. 0.00038;
Matches 43; Conservative 9; Mismatches 59; Indels 27; Gaps 6;

QY 15 TERSTAPLNTQISALRNDPKPLPQQPPAPAPANQDNSSQNTRLQPTPPI--PAPAPKPAAP 72
Db 78 TKAPTPPYKPPPTKA---PTP-PYKPPAPAPPTK-----APTPPYKPPAPAPPTKAP 124
QY 73 PRPLDRESPGVENKLIPSVGSPASSTPLPPDGTGPNSTPNNRVTPVSGSNSSSADPKA 132
Db 125 TPTYKPPAP-----APPTKAPTPP-FKPPAPAPPTKAPTPPYKPPPTPAPAPPVK 172
QY 133 PPPPPVSSGEPPTLGENP 150
Db 173 APTPPYMPPTPTKAPT 190

RESULT 9

US-08-819-013-1
; Sequence 1, Application US/08819013
; Patent No. 5994522
; GENERAL INFORMATION:
; APPLICANT: Chan, Andrew C.
; TITLE OF INVENTION: BLNK PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/819,013
; FILING DATE: 17-MAR-1997

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/788,322
; FILING DATE: 24-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-64383-1/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-819-013-1

Query Match 13.3%; Score 137; DB 2; Length 456;
Best Local Similarity 25.9%; Pred. No. 0.00083;
Matches 51; Conservative 26; Mismatches 58; Indels 62; Gaps 10;

QY 10 ISNNKTERSTAPLNTQISALRNDPKPLQOPPAPANQDNSSQNTRLQPTPIPAPAKP 69
Db 120 IDNRSSQRHSPPPS-----KTLPSKPSWP-----SEKARLTSTLPALTALQKP 162

QY 70 AAPRP---LDRESPGV-----ENKLIPSVGS--PASSTPLPPDGTGPNSTPNNAVT 117
Db 163 QVPPKPKGLLEADYVVPVEDNDENYIHTESSPPPEKAPMVRSTKPNSS-----T 216

QY 118 PVS-----QGSNSSSADPKAPPP---PPVSSGEPPTL-----GENPDGLSQEQ-- 157
Db 217 PASPPGTASGRNSGAWETKSPPPAAPSLPRAGKKPTTPLKTPVASQQNASVCEKPI 276

QY 158 -----LEHRERSLQT 167
Db 277 PAERHRGSSHROEAVQS 293

RESULT 10
US-09-252-991A-24873
; Sequence 24873, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24873
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24873

Query Match 13.3%; Score 136.5; DB 4; Length 142;
Best Local Similarity 26.8%; Pred. No. 0.00023;
Matches 41; Conservative 19; Mismatches 36; Indels 57; Gaps 7;

QY 7 IQNISNNKTERSTAPL-----NTQISALRNDPKPLQOPPAPANQDNSSQNTRLQPT 59
Db 25 VQAFCDSDSKPCTLSLPNGLFFFTNAQLSSLOPPCKPVP----- 62

QY 60 PPIPAPAKPAAPRPLDRSPGVENKLIPSVGSPASSTPLPPDGTGPNSTPNNAVT 119
```

```
Db 63 PPVPPSPSPSPSP--PVPPSPPV-----PGVPPSPSPSPVP--SPPSP----- 106
QY 120 SQGSNSSSADPKAPPPPPVSSGEPPTLGENPDG 152
Db 107 -----PPVPPSPSPS--PPPPVPPSPPG 127

RESULT 11
US-09-080-897-2
; Sequence 2, Application US/09080897
; Patent No. 5985574
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welcsh, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,897
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-080-897-2

Query Match 13.3%; Score 136.5; DB 2; Length 1248;
Best Local Similarity 32.3%; Pred. No. 0.0029;
Matches 43; Conservative 8; Mismatches 59; Indels 23; Gaps 5;

QY 35 PLPQQ-----PPAPANQDNSSQNTRLQPTPIPAPAKPA-----APRPLDRES----- 80
Db 578 PLPGDSGTIIPPPAPGD---STTPPPPPPPPPPPPLPGGTAISPPPLSGDATIPPP 633

QY 81 -PGVENKLIPSVGSPASSTPLPPDGTGPNSTPNNAVTVPVSGSNSSSADPKAPPPPPVS 139
Db 634 PPLPEGVGIPSPSSLPGGTAIPP-----PPPLPGSARIPPPPPPLPGSAGIPPPPLPGE 689

QY 140 SGEPTLGENPDG 152
Db 690 AGMPPPPPPLPG 702

RESULT 12
US-09-323-735-2
; Sequence 2, Application US/09323735
; Patent No. 6197932
```

RESULT 13
US-08-818-112-53
; Sequence 53, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

```

, , TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
, , TUBERCULOSIS
, ,
, , NUMBER OF SEQUENCES: 148
, ,
, , CORRESPONDENCE ADDRESS:
, , ADDRESSEE: SEED and BERRY LLP
, , STREET: 6300 Columbia Center, 701 Fifth Avenue
, , CITY: Seattle
, , STATE: Washington
, , COUNTRY: USA
, , ZIP: 98104-7092
, ,
, , COMPUTER READABLE FORM:
, , MEDIUM TYPE: Floppy disk
, ,

```

```
;
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-818-111-53

Query Match 13.1%; Score 135; DB 4; Length 332;
Best Local Similarity 27.6%; Pred. NO. 0.00084;
Matches 42; Conservative 15; Mismatches 57; Indels 38; Gaps 6;

QY 6 HIQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDQNSSQNTRLQPTPPIPAP 65
Db | : | | | | | : | : | : | | | | : | | | |
7 HMQVDPNLTRRKGRLAALAAMAS-ASLVTVAVPATANADP-----EPAPPVPTT 57

QY 66 APKP-----AAPRRPLDRESPGVENKLIPSVGSPASSTPLPPDGTGPNSTPNNRAVTPVSQ 121
Db | | | | | | | | : | | | | : | | | |

QY 122 GSNSSSADPKAPPPPPVSSGEP-PTLGENPDG 152
Db : | | | | | | | : | | | | : | | | |
94 NAAPPADPNAPPPPPVIAPNAPQPVRIIDNPVG 125

RESULT 15
US-09-056-556-53
; Sequence 53, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-09-056-556-53

Query Match 13.1%; Score 135; DB 4; Length 332;
Best Local Similarity 27.6%; Pred. NO. 0.00084;
Matches 42; Conservative 15; Mismatches 57; Indels 38; Gaps 6;

QY 6 HIQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDQNSSQNTRLQPTPPIPAP 65
Db | : | | | | | : | : | : | | | | : | | | |
7 HMQVDPNLTRRKGRLAALAAMAS-ASLVTVAVPATANADP-----EPAPPVPTT 57

QY 66 APKP-----AAPRRPLDRESPGVENKLIPSVGSPASSTPLPPDGTGPNSTPNNRAVTPVSQ 121
Db | | | | | | | | : | | | | : | | | |

QY 122 GSNSSSADPKAPPPPPVSSGEP-PTLGENPDG 152
Db : | | | | | | | : | | | | : | | | |
94 NAAPPADPNAPPPPPVIAPNAPQPVRIIDNPVG 125

Search completed: October 15, 2003, 10:32:01
Job time : 40.2529 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:32:06 ; Search time 82.2802 Seconds
(without alignments)
379.908 Million cell updates/sec

Title: US-09-915-543-15_COPY_199_392
Perfect score: 1028
Sequence: 1 TIVSFHQNISNNKTERSTA.....LFPDEKEFTGAQSGGPQONP 194

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*

17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1028	100.0	1426	15	US-10-322-579-15
2	178	17.3	35	15	US-10-322-579-5
3	167.5	16.3	707	11	US-09-919-039-278
4	162.5	15.8	574	12	US-10-168-097A-76
5	162.5	15.8	574	12	US-10-239-431A-38
6	158	15.4	802	9	US-09-823-240-2
7	154.5	15.0	731	14	US-10-086-464-17
8	152	14.8	509	12	US-10-021-660-86
9	151.5	14.7	503	14	US-10-078-547-2
10	150.5	14.6	533	11	US-09-922-226-2
11	150.5	14.6	533	15	US-10-098-184-2
12	150	14.6	5179	9	US-09-922-217-1068
13	150	14.6	5179	10	US-09-833-263-1068
14	150	14.6	5179	14	US-10-025-380-1068
15	149	14.5	507	14	US-10-078-547-24
					Sequence 15, Appl
					Sequence 5, Appl
					Sequence 278, Appl
					Sequence 76, Appl
					Sequence 38, Appl
					Sequence 17, Appl
					Sequence 86, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 2, Appl
					Sequence 1068, Ap
					Sequence 1068, Ap
					Sequence 24, Appl

16	148.5	14.4	731	14	US-10-086-464-8	Sequence 8, Appl
17	144.5	14.1	351	12	US-10-239-431A-22	Sequence 22, Appl
18	144.5	14.1	542	12	US-10-239-431A-10	Sequence 10, Appl
19	144.5	14.1	844	15	US-10-156-761-7663	Sequence 7663, Ap
20	142.5	13.9	721	14	US-10-086-464-5	Sequence 5, Appl
21	141.5	13.8	416	15	US-10-043-487-282	Sequence 282, App
22	141.5	13.8	505	12	US-10-168-097A-12	Sequence 12, Appl
23	141.5	13.8	505	12	US-10-239-431A-32	Sequence 32, Appl
24	141	13.7	666	9	US-09-791-171-70	Sequence 70, Appl
25	141	13.7	666	12	US-09-804-980-70	Sequence 70, Appl
26	141	13.7	666	16	US-10-080-170-640	Sequence 640, App
27	139.5	13.6	505	12	US-10-168-097A-56	Sequence 56, Appl
28	139.5	13.6	505	12	US-10-239-431A-36	Sequence 36, Appl
29	139	13.5	501	12	US-10-168-097A-46	Sequence 46, Appl
30	139	13.5	501	12	US-10-239-431A-35	Sequence 35, Appl
31	139	13.5	638	15	US-10-038-010-4	Sequence 4, Appl
32	138	13.4	377	15	US-10-149-819-7	Sequence 7, Appl
33	137.5	13.4	647	14	US-10-086-464-2	Sequence 2, Appl
34	137.5	13.4	647	14	US-10-086-464-4	Sequence 4, Appl
35	136.5	13.3	523	12	US-10-017-161-1982	Sequence 1982, Ap
36	136	13.2	633	12	US-10-168-097A-66	Sequence 66, Appl
37	136	13.2	633	12	US-10-239-431A-37	Sequence 37, Appl
38	135	13.1	332	12	US-10-084-843-53	Sequence 53, Appl
39	135	13.1	332	12	US-10-193-002-53	Sequence 53, Appl
40	135	13.1	332	12	US-10-098-732A-41	Sequence 41, Appl
41	134.5	13.1	200	9	US-09-925-297-700	Sequence 700, App
42	134.5	13.1	668	10	US-09-925-300-1633	Sequence 1633, Ap
43	134.5	13.1	668	12	US-10-240-154-20	Sequence 20, Appl
44	134.5	13.1	704	12	US-10-240-154-18	Sequence 18, Appl
45	134	13.0	1228	11	US-09-917-384-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-10-322-579-15

; Sequence 15, Application US/10322579

; Publication No. US20030114413A1

; GENERAL INFORMATION:

; APPLICANT: BASLER, Konrad

; APPLICANT: BRUNNER, Erich

; APPLICANT: FROESCH, Barbara

; APPLICANT: KRAMPS, Thomas

; APPLICANT: PETER, Oliver

; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATH

; THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON

; FILE REFERENCE: Q60361

; CURRENT APPLICATION NUMBER: US/10/322,579

; CURRENT FILING DATE: 2002-12-19

; PRIOR APPLICATION NUMBER: US/09/915,543

; PRIOR FILING DATE: 2001-07-27

; PRIOR APPLICATION NUMBER: 60/221,502

; PRIOR FILING DATE: 2000-07-28

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 15

; LENGTH: 1426

; TYPE: PRT

; ORGANISM: Human lgs/bcl9

US-10-322-579-15

Query Match	100.0%;	Score 1028;	DB 15;	Length 1426;
Best Local Similarity	100.0%;	Pred. No. 7.9e-59;		
Matches	194;	Conservative	0;	Mismatches 0;
		Indels	0;	Gaps 0;
QY	1	TIVSFHQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNRLQPTP	60	
Db	199	TIVSFHQNISNNKTERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNRLQPTP	258	
QY	61	PIPAPAKPAAPRPLDRSPGVENKLIPSVGSPASSTPLPPDGTGPNSTPNNRAVTPVS	120	
Db	259	PIPAPAKPAAPRPLDRSPGVENKLIPSVGSPASSTPLPPDGTGPNSTPNNRAVTPVS	318	

Db 153 APPPPRRSRAGSAGALPCAGHTRRRTSSPRSSPPLSGPPGRASPRGAPPPPLLRAA 212
QY 110 -TPNNRAVTPVSGSNSSADPKAPPPPPVSSGE-----PP-----TL 146
Db 213 PTPSPRALAP-----AAASPPPPPPPPGREGKRRKKFFPGSSGSTQTSGAAAVAAAL 265
QY 147 GENPDGLSQEQLHERSLQTLRDIQRLMFPDEKEFTGAQSG 188
Db 266 GSSPG-----RRLLPLLVRG-----PRSGAASG 291

RESULT 9
US-10-078-547-2
; Sequence 2, Application US/10078547
; Publication No. US20020199211A1
; GENERAL INFORMATION:
; APPLICANT: Narayanaswamy Ramesh
; APPLICANT: Miguel A. de la Fuente
; APPLICANT: Ines M. Anton
; APPLICANT: Raif S. Geha
; TITLE OF INVENTION: WIP, A WASP-Associated Protein
; FILE REFERENCE: 1242.1022-005
; CURRENT APPLICATION NUMBER: US/10/078,547
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/599,287
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/068,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Human
US-10-078-547-2

Query Match 14.7%; Score 151.5; DB 14; Length 503;
Best Local Similarity 26.5%; Pred. No. 0.019;
Matches 60; Conservative 25; Mismatches 84; Indels 57; Gaps 11;
QY 12 NNK-----TERSTAPLNTQISALRNDPKLPQQPAPANQDQNS-----SQNTRL-QPT 59
Db 289 NNKPPVSTPRPSAPHRPH---LRPPPSRPGPPPLPPSSGNDETRLPQRLSLSSST 345
QY 60 PPIAPAPKPAAPRPLDRESFGVENKLIPIVSGSPASSTPLPPDGTGPNSTNNRAV--- 116
Db 346 PPLPSFGSGPLPPPPSERPPPPVRD-----PPGRSGPLPPPPPPVSRNGSTSRALPAT 398
QY 117 --TPVSGSNSSADPK-----APPPPVSSGEPPTLGENPDGLSQEQLHR-- 161
Db 399 PQLPSRSGVDSRSGPRPLPDRPSAGAPPPPPSTSRNGFQDSP---CEDEWESRFY 455
QY 162 -----ERSLQTLRDIQRLMFPDEKEFTGA---QSGGQQNP 194
Db 456 FHPISDLPPPEYVQTTKSPSKLARNESR-SGSNNRERGPPPLPP 500

RESULT 10
US-09-922-226-2
; Sequence 2, Application US/09922226
; Publication No. US20030077664A1
; GENERAL INFORMATION:
; APPLICANT: Zhao, Yi
; APPLICANT: Thacher, Scott M.
; APPLICANT: Xiao, Jia-Hao
; APPLICANT: Kusari, Jyotirmoy
; APPLICANT: Chandraratna, Roshantha A.
; TITLE OF INVENTION: Methods of Screening For Compounds That
; TITLE OF INVENTION: Modulate Hormone Receptor Activity

FILE REFERENCE: P-AR 4681
CURRENT APPLICATION NUMBER: US/09/922,226
CURRENT FILING DATE: 2002-01-09
PRIOR APPLICATION NUMBER: US 60/284,797
PRIOR FILING DATE: 2001-04-18
NUMBER OF SEQ ID NOS: 191
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 533
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-226-2
Query Match 14.6%; Score 150.5; DB 11; Length 533;
Best Local Similarity 31.0%; Pred. No. 0.023;
Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;
QY 31 NDPKPLPQ--QPPAPANQDNSSQNTRLQPTTPIPAPA-----PKPAAPRPLDRESP 81
Db 86 SSNPPLPQGVPPSPPG-----PPLPSTAPSLGGSGAGPPPPMPPPLGSPFP 134
QY 82 GVENKLIPSVGSPASSTPLPPDGTGPNSTNNRAVTPVSGSNSSADPKAP----- 133
Db 135 VISS----SMGSPGLPPAPPGFSGPVSSPQINSTVSLPGGSGPPEDVKPPVLGVRGLH 190
QY 134 -PPPPVSSG 141
Db 191 CPPPPGGPG 199
RESULT 11
US-10-098-184-2
; Sequence 2, Application US/10098184
; Publication No. US20030105333A1
; GENERAL INFORMATION:
; APPLICANT: Pfahl, Magnus
; APPLICANT: Tachdjian, Catherine
; APPLICANT: Al-Shamma, Hussien A.
; APPLICANT: Fanjul, Andrea
; APPLICANT: Pleyner, David P.M.
; APPLICANT: Spruce, Lyle W.
; APPLICANT: Fine, Richard
; APPLICANT: Zapf, James W.
; TITLE OF INVENTION: RXR ACTIVATING MOLECULES
; FILE REFERENCE: 13099.0016U2
; CURRENT APPLICATION NUMBER: US/10/098,184
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: 60/274,342
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 533
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. US20030105333A1e -
; OTHER INFORMATION: synthetic construct
US-10-098-184-2
Query Match 14.6%; Score 150.5; DB 15; Length 533;
Best Local Similarity 31.0%; Pred. No. 0.023;
Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;
QY 31 NDPKPLPQ--QPPAPANQDNSSQNTRLQPTTPIPAPA-----PKPAAPRPLDRESP 81
Db 86 SSNPPLPQGVPPSPPG-----PPLPSTAPSLGGSGAGPPPPMPPPLGSPFP 134
QY 82 GVENKLIPSVGSPASSTPLPPDGTGPNSTNNRAVTPVSGSNSSADPKAP----- 133
Db 135 VISS----SMGSPGLPPAPPGFSGPVSSPQINSTVSLPGGSGPPEDVKPPVLGVRGLH 190
QY 134 -PPPPVSSG 141

```
Db      191 CPPPPGGPG 199
||||| |
RESULT 12
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068

Query Match      14.6%; Score 150; DB 9; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.24;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY      15 TERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQTPPIAPAPAKPAAP-- 72
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1428 TTTTPPPTTTPSPPIITTTTTLPLTTTPSPPISTITTPPPT-TTPSPPTTTPSPPTTTPSP 1486

QY      73 -----PRPLDRESPGVENKLIPSVGSPASSTPLPPDGT-----GPNSTPNR 114
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1487 PTTTTPPPTTTPSPPMTPITP-----PASTTTLPPPTTTPSPPTTTTTPPPTTTPSP 1542

QY      115 AVTPVSGSNSSADPKAPP-----PPPVSSGEPPT 145
| | : : : : | | | | : : | |
Db      1543 TTTTTPPTSTTTLPPPTTTPSPPTTTTTPPPTTTPSPPT 1582
| | : : : : | | | | : : | |

; ORGANISM: Homo sapiens
US-09-922-217-1068

Query Match      14.6%; Score 150; DB 9; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.24;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY      15 TERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQTPPIAPAPAKPAAP-- 72
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1428 TTTTPPPTTTPSPPIITTTTTLPLTTTPSPPISTITTPPPT-TTPSPPTTTPSPPTTTPSP 1486

QY      73 -----PRPLDRESPGVENKLIPSVGSPASSTPLPPDGT-----GPNSTPNR 114
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1487 PTTTTPPPTTTPSPPMTPITP-----PASTTTLPPPTTTPSPPTTTTTPPPTTTPSP 1542

QY      115 AVTPVSGSNSSADPKAPP-----PPPVSSGEPPT 145
| | : : : : | | | | : : | |
Db      1543 TTTTTPPTSTTTLPPPTTTPSPPTTTTTPPPTTTPSPPT 1582
| | : : : : | | | | : : | |

; ORGANISM: Homo sapiens
US-09-833-263-1068

Query Match      14.6%; Score 150; DB 14; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.24;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY      15 TERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQTPPIAPAPAKPAAP-- 72
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1428 TTTTPPPTTTPSPPIITTTTTLPLTTTPSPPISTITTPPPT-TTPSPPTTTPSPPTTTPSP 1486

QY      73 -----PRPLDRESPGVENKLIPSVGSPASSTPLPPDGT-----GPNSTPNR 114
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1487 PTTTTPPPTTTPSPPMTPITP-----PASTTTLPPPTTTPSPPTTTTTPPPTTTPSP 1542

QY      115 AVTPVSGSNSSADPKAPP-----PPPVSSGEPPT 145
| | : : : : | | | | : : | |
Db      1543 TTTTTPPTSTTTLPPPTTTPSPPTTTTTPPPTTTPSPPT 1582
| | : : : : | | | | : : | |

; ORGANISM: Homo sapiens
US-09-833-263-1068
```

```
Query Match      14.6%; Score 150; DB 10; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.24;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY      15 TERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQTPPIAPAPAKPAAP-- 72
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1428 TTTTPPPTTTPSPPIITTTTTLPLTTTPSPPISTITTPPPT-TTPSPPTTTPSPPTTTPSP 1486

QY      73 -----PRPLDRESPGVENKLIPSVGSPASSTPLPPDGT-----GPNSTPNR 114
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1487 PTTTTPPPTTTPSPPMTPITP-----PASTTTLPPPTTTPSPPTTTTTPPPTTTPSP 1542

QY      115 AVTPVSGSNSSADPKAPP-----PPPVSSGEPPT 145
| | : : : : | | | | : : | |
Db      1543 TTTTTPPTSTTTLPPPTTTPSPPTTTTTPPPTTTPSPPT 1582
| | : : : : | | | | : : | |

; ORGANISM: Homo sapiens
US-10-025-380-1068

; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1068

Query Match      14.6%; Score 150; DB 14; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.24;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY      15 TERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQTPPIAPAPAKPAAP-- 72
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1428 TTTTPPPTTTPSPPIITTTTTLPLTTTPSPPISTITTPPPT-TTPSPPTTTPSPPTTTPSP 1486

QY      73 -----PRPLDRESPGVENKLIPSVGSPASSTPLPPDGT-----GPNSTPNR 114
| : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1487 PTTTTPPPTTTPSPPMTPITP-----PASTTTLPPPTTTPSPPTTTTTPPPTTTPSP 1542

QY      115 AVTPVSGSNSSADPKAPP-----PPPVSSGEPPT 145
| | : : : : | | | | : : | |
Db      1543 TTTTTPPTSTTTLPPPTTTPSPPTTTTTPPPTTTPSPPT 1582
| | : : : : | | | | : : | |

; ORGANISM: Homo sapiens
US-10-078-547-24
; Sequence 24, Application US/10078547
; Publication No. US20020199211A1
; GENERAL INFORMATION:
```

Wed Oct 15 11:40:25 2003

us-09-915-543-15_copy_199_392.rapb

Page 6

```

; APPLICANT: Narayanaswamy Ramesh
; APPLICANT: Miguel A. de la Fuente
; APPLICANT: Ines M. Anton
; APPLICANT: Raif S. Geha
; TITLE OF INVENTION: WIP, A WASP-Associated Protein
; FILE REFERENCE: 1242.1022-005
; CURRENT APPLICATION NUMBER: US/10/078,547
; CURRENT FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: 09/599,287
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: PCT/US98/27501
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/101,457
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/068,533
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translated WIP ORF No. US20020
US-10-078-547-24

```

Query Match	14.5%;	Score 149;	DB 14;	Length 507;
Best Local Similarity	26.9%;	Pred. No. 0.027;		
Matches 61;	Conservative 25;	Mismatches 83;	Indels 58;	Gaps 12;
QY	12	NNK-----TERSTAPLNTQISALRNDPKPLPQQPPAPANQDQNS-----SONTRL-QPT	59	
Db	292	NNKPPVPVSTPRPSAPHRPH--LRPPPPSRPGPPPLPPSSSGNDETPrLPQRLNLSST	348	
QY	60	PPIPAPAK-PAAPRPLDRESpgVENKLIPSVGSPASSTPLPDDGTGPNSTPNNRAV--	116	
Db	349	PPLPSGRSGPLPPPVPSPERPppVRD-----PGRSGPLPPPPVSRNGSTSRALPA	401	
QY	117	---TPVSGSNSSSADPK-----APPPPPVSSGEPTLGENPDGLSQELEHR-	161	
Db	402	TPQLPSRGVDSRSGRPPLPDDRPSAGAPPPPPSTSRNGFQDSP---CEDEWESRF	458	
QY	162	-----ERSLQTLRDIQRMFPDEKEFTGA---QSGGPPQNP	194	
Db	459	YFHPISDLPPPEPVQTTKSPSKLARNSR-SGSNNRRERGPPPLPP	504	

Search completed: October 15, 2003, 10:50:59
Job time : 83.2802 secs

Query Match 16.4%; Score 168.5; DB 2; Length 876;
Best Local Similarity 30.1%; Pred. No. 0.012;
Matches 63; Conservative 24; Mismatches 61; Indels 61; Gaps 10;
QY 3 VSFHIQINISNKT---ERS--TAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQ 57
Db 177 VSLNDNSTIGDKTGNEKSLDTAPSSSTKKLSISSIASALPAMPWSPSATDE-----AV 229
QY 58 PTPPIAPAPKPAAPRPLDRESPGVENKLIPSVGSPASSTPLPDGTGPNSTPNNAV 117
Db 230 KSPPAASAAAAAAPPAP-----APAAALTPAPP-----PPVTRKLT 265
QY 118 -PVSQGSNSSADPKAPPPPVSSGEP-----PTLGENPDG-----LSQE----- 156
Db 266 SPFSWLSRSSKEQANPPPPPPAAATAPRRNTASSVATLSSNPDGLGKVDEEGNKNTLKD 325
QY 157 ---QLEHRER-SLQTLRDIQRMFLFPDEKE 181
Db 326 RFKQLRLREGSAPTGDDEKSTPEEKE 354

RESULT 3
A46302
PTB-associated splicing factor, long form - human
N;Alternate names: myoblast cell surface antigen 24.1D5; polypyrimidine tract-binding pr
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 24-Sep-1999
C;Accession: A46302; A43557; S2995
R;Patton, J.G.; Porro, E.B.; Galceran, J.; Tempst, P.; Nadal-Ginard, B.
Genes Dev. 7, 393-406, 1993
A;Title: Cloning and characterization of PSF, a novel pre-mRNA splicing factor.
A;Reference number: A46302; MUID:93194059; PMID:8449401
A;Accession: A46302
A;Molecule type: mRNA
A;Residues: 1-707 <PAT>
A;Cross-references: EMBL:X70944; NID:g38457; PIDN:CAA50283.1; PID:g38458
A;Note: sequence extracted from NCBI backbone (NCBIP:127206)
R;Gower, H.J.; Moore, S.E.; Dickson, G.; Elsom, V.L.; Nayak, R.; Walsh, F.S.
Development 105, 723-731, 1989
A;Title: Cloning and characterization of a myoblast cell surface antigen defined by 24.1
A;Reference number: A43557; MUID:90091812; PMID:2480877
A;Accession: A43557
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 312-707 <GOW>
A;Cross-references: GB:X16850; NID:g23711; PIDN:CAA34747.1; PID:g23712
C;Genetics:
A;Gene: GDB:SFPO; PSF
A;Cross-references: GDB:138275
A;Map position: 4q-4q
C;Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein
C;Keywords: alternative splicing; pre-mRNA splicing; surface antigen
F;298-359/Domain: ribonucleoprotein repeat homology <RRM1>
F;372-438/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 16.3%; Score 167.5; DB 2; Length 707;
Best Local Similarity 40.0%; Pred. No. 0.011;
Matches 54; Conservative 5; Mismatches 47; Indels 29; Gaps 8;
QY 30 RNDPKPLPQQPPAPANQDN--SSQNTRLQPTPIAPAPAKP-----AAPRPLDRESPGV 83
Db 53 QSGPKP-PIPPPPHQQQQPPPPQPPPPHQPHPHQPQQQQPPPPQDSSKPVV 111
QY 84 EN--KLIPSVGS--PASSTPLP-----PDGTGPNSTPN-NRAVTPVSGSNSSADP 130
Db 112 AQGPGPAPGVGSAPPASSAPPATPTISGAPPGSGPGTPTTPPAVT-----SAPP 162
QY 131 KAPPPPPVSSGEPPT 145
Db 163 GAPPTTPSSGVPTT 177

RESULT 4
T43556
Wiskott-Aldrich syndrome protein homolog - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T43556
R;Zankel, T.C.; Ow, D.W.
submitted to the EMBL Data Library, December 1997
A;Description: A Wiskott-Aldrich Syndrome protein homolog in Schizosaccharomyces pombe
A;Reference number: Z22575
A;Accession: T43556
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-574 <ZAN>
A;Cross-references: EMBL:AF038575; PIDN:AAB92587.1
A;Experimental source: strain JS21
C;Genetics:
A;Gene: wsp1
A;Map position: 1
A;Introns: 72/3; 519/3; 564/1

Query Match 15.8%; Score 162.5; DB 2; Length 574;
Best Local Similarity 30.9%; Pred. No. 0.017;
Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;
QY 7 IQNISNNKTERSPA-----PLNTQISA---LRNDPKPLPQQPPAPANQDQ-----NSSQNT 54
Db 275 IAPVSMNPAINSTSKPPLPPPPSSRVSAALAANKRPPPPPPSRNRGRKPPIGNGSSNS 334
QY 55 RLQPTPP-----IPAPAKPAAPRPLDRESPGVENKLIPSVGSPA-SSTPLPPDG 104
Db 335 SLPPPPPPPPRSNAAGSIPLPQQRSAPPPPPPSAPSTGRQPPPLSSSRVSNPPAPPPA 394
QY 105 TGPNSTPNNRAVTPVSGSNSSSSADPKAPPPPPVSSGEPPTL 146
Db 395 IGRSAP---ALPPLGNASRTST--PPVTPPSSLPPSAPPSL 431

RESULT 5
T38819
wiskott-aldrich syndrome protein homolog 1 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38819
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21813
A;Accession: T38819
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-574 <CON>
A;Cross-references: EMBL:Z98980; NID:e1060691; PIDN:CAB11718.1; GSPDB:GN000066; SPDB:S
A;Experimental source: strain 972h-; cosmid c4F10
C;Genetics:
A;Gene: wsp1; SPDB:SPAC4f10.15c
A;Map position: 1
A;Introns: 72/3; 519/3; 564/1

Query Match 15.8%; Score 162.5; DB 2; Length 574;
Best Local Similarity 30.9%; Pred. No. 0.017;
Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;
QY 7 IQNISNNKTERSTA-----PLNTQISA---LRNDPKPLPQQPPAPANQDQ-----NSSQNT 54
Db 275 IAPVSMNPAINSTSKPPLPPPPSSRVSAALAANKRPPPPPPSRNRGRKPPIGNGSSNS 334
QY 55 RLQPTPP-----IPAPAKPAAPRPLDRESPGVENKLIPSVGSPA-SSTPLPPDG 104
Db 335 SLPPPPPPPPRSNAAGSIPLPQQRSAPPPPPPSAPSTGRQPPPLSSSRVSNPPAPPPA 394
QY 105 TGPNSTPNNRAVTPVSGSNSSSSADPKAPPPPPVSSGEPPTL 146
Db 395 IGRSAP---ALPPLGNASRTST--PPVTPPSSLPPSAPPSL 431

S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Query Match 15.4%; Score 158.5; DB 2; Length 1188;
Best Local Similarity 32.0%; Pred. No. 0.057;
Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;

QY 14 KTERSTAPLNTQISALRNDPKPLP--QQPPAPANQDNSSQNTRLQP---TPPIPAPAPK 68
| | | | | : : : | | | | | : : : | | | | |
Db 961 KSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKSSPPAPMSSPPPEVKSPPPAPVSS 1020

QY 69 P-----AAPRPLDRESGVENKLIPSVGSPASSTPLP---PDGTGPNSTNNRAVTPV 119
| | | | | : : : | | | | | : : : | | | | |
Db 1021 PPPPVKSPPPAPVSSPPPPVKSS---PPPPAPVSSPPPPVKSPPPAPISSPPPPVKSP 1077

QY 120 SQGSNSSADP-KAPPPP-PVSSGEPPTLGENP 150
| | | | | : : : | | | | | : : : | | | | |
Db 1078 PPAPVSSPPPPVKSPPPAPVSSPPPIKSPPP 1110

RESULT 11
S22697
extensin - Volvox carteri (fragment)
C:Species: Volvox carteri
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 21-Jul-2000
C:Accession: S22697; S21006
R:Ertl, H.; Hallmann, A.; Wenzl, S.; Sumper, M.
EMBO J. 11, 2055-2062, 1992
A:Title: A novel extensin that may organize extracellular matrix biogenesis in Volvox ca
A:Reference number: S22697; MUID:92289669; PMID:1600938
A:Accession: S22697
A:Molecule type: mRNA
A:Residues: 1-464 <HAL>
A:Cross-references: EMBL:X65165; NID:g21991; PIDN:CAA46283.1; PID:g21992
C:Keywords: glycoprotein

Query Match 15.3%; Score 157.5; DB 2; Length 464;
Best Local Similarity 31.3%; Pred. No. 0.026;
Matches 41; Conservative 11; Mismatches 64; Indels 15; Gaps 3;

QY 21 PLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQTPPIPAPAPKPAAPRPLDRES 80
| | | | | : : : | | | | | : : : | | | | |
Db 292 PPPPRVSPPPPPQPVSSPPPPPPPPRSPSPPPRSPSPSPPPSPSPSPSPSPSPSPSP 351

QY 81 PGVENKLIPSVGSPASSTPLPPDGTGPNSTNNRAVTPVSGSNSSSADPKAPPP-PPVS 139
| | | | | : : : | | | | | : : : | | | | |
Db 352 P---PRSSPSPPPPVWSPPPPPPRASPPPPP-----ASSPPPPPPPPPPSPPPS 397

QY 140 SGEPTTIGENP 150
| | | | |
Db 398 PPPPATAAANP 408

RESULT 12
S16748
proline-rich protein - rape (fragment)
C:Species: Brassica napus (rape)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C:Accession: S16748
R:Roberts, M.R.; Foster, G.D.; Blundell, R.P.; Robinson, S.W.; Draper, J.; Scott, R.J.
submitted to the EMBL Data Library, August 1991

A:Description: Cloning and characterization of a proline-rich gene expressed specifi
A:Reference number: S16748
A:Accession: S16748
A:Molecule type: mRNA
A:Residues: 1-449 <ROB>
A:Cross-references: EMBL:X60376; NID:g22596; PID:g22597

Query Match 15.2%; Score 156; DB 2; Length 449;
Best Local Similarity 33.3%; Pred. No. 0.03;
Matches 39; Conservative 9; Mismatches 57; Indels 12; Gaps 3;

QY 33 PKPLPQQPPAPANQDNSSQNTRLQTPPIPAPAPKPAAPRPLDRESGVENKLIPSVG 92
| | | | | : : : | | | | | : : : | | | | |
Db 2 PKPQPKPPPKPQPKPPAPTSPCPPQPKPQPKPPAPTSPCPPQPKPQPKPPAPG 61

QY 93 -----SPASSTPLPPDGTGPNSTNNRAVTPVSGSNSSSADPKAPPPPPVSSGEP 144
| | | | | : : : | | | | | : : : | | | | |
Db 62 PSPKPGSPSPKPPSPAPKVP-----PPSPSPKPSPPKPPAPSPKP-SPPKPP 111

RESULT 13
T04455
hypothetical protein F4D11.90 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 04-Mar-2000
C:Accession: T04455
R:Bevan, M.; Benes, V.; Rechmann, S.; Borkova, D.; Anson, W.; Hoheisel, J.; Mewes,
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15360
A:Accession: T04455
A:Molecule type: DNA
A:Residues: 1-731 <BEV>
A:Cross-references: EMBL:AL022537
A:Experimental source: cultivar Columbia; BAC clone F4D11
C:Genetics:
A:Map position: 4
A:Introns: 326/1; 476/1; 505/1; 528/3; 557/2; 606/3; 660/3
A:Note: F4D11.90
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hon

Query Match 15.0%; Score 154.5; DB 2; Length 731;
Best Local Similarity 27.6%; Pred. No. 0.06;
Matches 47; Conservative 20; Mismatches 68; Indels 35; Gaps 7;

QY 15 TERSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTRLQTPPIPAPAPK--PAAP 72
| | | | | : : : | | | | | : : : | | | | |
Db 29 TSSPPAPPLSLPLPPPLSSPPPLSPPLSAPTASPPPLPVESPPSPPIESPPPPLESPP 88

QY 73 PRPLDRESGVENKLIPSVGSPASSTPL-----PPDGTGPNST-PNNRAVTPVSG- 122
| | | | | : : : | | | | | : : : | | | | |
Db 89 PPPIESPPSP-----SPHVSAPSGSPPLFLPAKPSPPSPSPSETVPPGNTISPPRSL 143

QY 123 -----SNSSSADPKAPP-----PPPVSSGEPPTLGENPDGLSQ 155
| | | | | : : : | | | | | : : : | | | | |
Db 144 PSESTPPVNTASPPPPSPRRRRSGPKSPFPPIINS-SPPNPSNTPSLPE 192

RESULT 14
T05441
proline-rich protein F7K2.50 - Arabidopsis thaliana (fragment)
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 26-Aug-1999
C:Accession: T05441
R:Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue
submitted to the Protein Sequence Database, November 1998
A:Reference number: Z15416
A:Accession: T05441
A:Molecule type: DNA
A:Residues: 1-379 <BEV>
A:Cross-references: EMBL:AL033545
A:Experimental source: cultivar Columbia; BAC clone F7K2
C:Genetics:
A:Map position: 4

```
A>Note: F7K2.50
C;Superfamily: hydroxyproline-rich glycoprotein

Query Match      14.8%; Score 152.5; DB 2; Length 379;
Best Local Similarity 31.8%; Pred. No. 0.041;
Matches 41; Conservative 12; Mismatches 45; Indels 31; Gaps 6;

QY   33 PKPLPQQPPAPANQDNSSQNTRLQPTTPI---PAPAPKPAAPRRLDRESGCVENKLIP 89
     | :| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   51 PQDPQPPTTPPT-----FQAPPANDQPPPPQSTSP--PVATTPPALPKPLP 98

QY   90 SVGSPASSTLPDPGTGNSTPNNRVTPVSQGSNSSSADPKAPPPPPVSSGEPTLGEN 149
     || :||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   99 PPLSPQTTPPPPAITPPPPP---AITP-----PLSPPPPAII--PPPLATT 142

QY   150 PDGLSQEQL 158
     | | | : |
Db   143 PPALPKPL 151

RESULT 15
D85257
extensin-like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C;Accession: D85257
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: D85257
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-379 <STO>
A;Cross-references: GB:NC_001268; NID:g7269093; PIDN:CAB79202.1; GSPDB:GN00140
C;Genetics:
A;Gene: AT4g22470
A;Map position: 4
C;Superfamily: hydroxyproline-rich glycoprotein

Query Match      14.8%; Score 152.5; DB 2; Length 379;
Best Local Similarity 31.8%; Pred. No. 0.041;
Matches 41; Conservative 12; Mismatches 45; Indels 31; Gaps 6;

QY   33 PKPLPQQPPAPANQDNSSQNTRLQPTTPI---PAPAPKPAAPRRLDRESGCVENKLIP 89
     | :| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   51 PQDPQPPTTPPT-----FQAPPANDQPPPPQSTSP--PVATTPPALPKPLP 98

QY   90 SVGSPASSTLPDPGTGNSTPNNRVTPVSQGSNSSSADPKAPPPPPVSSGEPTLGEN 149
     || :||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   99 PPLSPQTTPPPPAITPPPPP---AITP-----PLSPPPPAII--PPPLATT 142

QY   150 PDGLSQEQL 158
     | | | : |
Db   143 PPALPKPL 151

Search completed: October 15, 2003, 10:33:12
Job time : 48.0467 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 23.4008 Seconds
(without alignments)
389.867 Million cell updates/sec

Title: US-09-915-543-15_COPY_199_392

Perfect score: 1028

Sequence: 1 TIVSFHIQINISNNKTERSTA.....LFPDEKEFTGAQSGGPGQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	1024	99.6	1426	1 BCL9_HUMAN	O00512 homo sapien
2	167.5	16.3	707	1 SFPQ_HUMAN	P23246 homo sapien
3	161.5	15.7	815	1 MK07_HUMAN	Q13164 homo sapien
4	159.5	15.5	555	1 GPl_CHLRE	Q9fpg6 chlamydomon
5	158	15.4	802	1 ENAH_MOUSE	Q03173 mus musculu
6	156	15.2	449	1 APG_BRANA	P40603 brassica na
7	151.5	14.7	503	1 WAIP_HUMAN	O43516 homo sapien
8	150.5	14.6	251	1 PRP2_HUMAN	P02812 homo sapien
9	150.5	14.6	533	1 RXRB_HUMAN	P28702 homo sapien
10	150	14.6	5179	1 MUC2_HUMAN	Q02817 homo sapien
11	149	14.5	426	1 EXLP_TOBAC	Q03211 nicotiana t
12	147	14.3	347	1 CSP_PLABA	P23093 plasmodium
13	146	14.2	534	1 APG_ARATH	P40602 arabidopsis
14	146	14.2	2167	1 SHK1_RAT	Q9wv48 rattus norv
15	144.5	14.1	339	1 CSP_PLABE	P06915 plasmodium
16	144.5	14.1	542	1 ZYX_CHICK	Q04584 gallus gall
17	144	14.0	535	1 SPKC_SYNY3	P74745 synecocyst
18	144	14.0	817	1 VRP1_YEAST	P37370 saccharomyc
19	143.5	14.0	520	1 RXRB_MOUSE	P28704 mus musculu
20	143	13.9	296	1 PRP3_MOUSE	P05143 mus musculu
21	143	13.9	1206	1 FM14_MOUSE	Q05859 mus musculu
22	143	13.9	1468	1 FMN1_MOUSE	Q05860 mus musculu
23	142.5	13.9	276	1 PRPL_HUMAN	P10162 homo sapien
24	141.5	13.8	283	1 EXTN_SORBI	P24152 sorghum bic
25	141.5	13.8	505	1 WASL_HUMAN	O00401 homo sapien
26	141.5	13.8	620	1 EXTN_TOBAC	P13983 nicotiana t
27	140.5	13.7	582	1 MNT_HUMAN	Q99583 homo sapien
28	140	13.6	806	1 MK07_MOUSE	Q9wvs8 mus musculu
29	139.5	13.6	268	1 NO20_MEDTR	P93329 medicago tr
30	139.5	13.6	505	1 WASL_BOVIN	Q95107 bos taurus
31	139.5	13.6	1664	1 SLP1_CLOTM	Q06852 clostridium
32	139	13.5	501	1 WASL_RAT	O08816 rattus norv
33	138.5	13.5	2161	1 SHK1_HUMAN	Q9y566 homo sapien

ALIGNMENTS

RESULT 1

ID	BCL9_HUMAN	STANDARD;	PRT;	1426 AA.
AC	O00512;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).			
GN	BCL9.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=98158621; PubMed=9490669;			
RA	Willis T.G., Zalcberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,			
RA	Jadavayel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,			
RA	Dyer M.J.S.;			
RT	"Molecular cloning of translocation t(1;14)(q21;q32) defines a novel			
RT	gene (BCL9) at chromosome 1q21.";			
RL	Blood 91:1873-1881(1998).			
RN	[2]			
RP	FUNCTION.			
RX	MEDLINE=21952490; PubMed=11955446;			
RA	Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,			
RA	Murone M., Zuellig S., Basler K.;			
RT	"Wnt/wingless signaling requires BCL9/legless-mediated recruitment of			
RT	pygopus to the nuclear beta-catenin-TCF complex.";			
RL	Cell 109:47-60(2002).			
CC	!- FUNCTION: Involved in signal transduction through the wnt pathway.			
CC	!- SUBUNIT: Binds to beta-catenin (CTNBL1), PYGO1 and PYGO2.			
CC	!- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,			
CC	testis, ovary and small intestine, and at lower levels in spleen,			
CC	colon and blood.			
CC	!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation			
CC	found in a patient with precursor B-cell acute lymphoblastic			
CC	leukemia (ALL). This translocation leaves the coding region			
CC	intact, but may have pathogenic effects due to alterations in the			
CC	expression level of BCL9. Several cases of translocations within			
CC	the 3' untranslated region of BCL9 have been found in B-cell			
CC	malignancies.			
CC	!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.			
CC	!- CAUTION: Ref.1 sequence differs from that shown due to a			
CC	frameshift in position 1391.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			

P10220 herpes simp
P10161 homo sapien
P05142 mus musculu
Q10341 orgyia pseu
P03330 simian sarc
O83384 treponema p
P49743 rattus norv
O60610 homo sapien
O02626 caenorhabdi
P48634 homo sapien
Q12446 saccharomyc
O88935 mus musculu

34 138.5 13.5 3164 1 TEGU_HSV11
35 137.5 13.4 234 1 PRPM_HUMAN
36 137 13.3 261 1 PRP2_MOUSE
37 137 13.3 279 1 Y091_NPVOP
38 137 13.3 512 1 GAG_SMSAV
39 137 13.3 516 1 Y369_TREPA
40 136.5 13.3 458 1 RXRB_RAT
41 136.5 13.3 1248 1 DIAL_HUMAN
42 136.5 13.3 1409 1 AEX3_CAEEL
43 136.5 13.3 2142 1 BAT2_HUMAN
44 136 13.2 633 1 LA17_YEAST
45 136 13.2 670 1 SYN1_MOUSE

QY 84 EN--KLIPSVGS--PASSTPLP-----PDGTGPNSTPN-NRAVTPVSQGSNSSADP 130
Db 112 AQGPGAPGVGSAPPASSAPPATPTTSGAPPGSGGPTTTPPPAVT-----SAPP 162
QY 131 KAPPPPPVSSGEPPT 145
Db 163 GAPPTTPSSGVPTT 177

RESULT 3
MK07_HUMAN
ID MK07_HUMAN STANDARD; PRT; 815 AA.
AC Q13164; Q16634;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mitogen-activated protein kinase 7 (EC 2.7.1.-) (Extracellular signal-
regulated kinase 5) (ERK-5) (ERK4) (BMK1 kinase).
GN MAPK7 OR PRKM7 OR ERK5 OR ERK4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95279403; PubMed=7759517;
RA Zhou G., Bao Z.Q., Dixon J.E.;
RT "Components of a new human protein kinase signal transduction
pathway.";
RL J. Biol. Chem. 270:12665-12669(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=95374539; PubMed=7646528;
RA Lee J.-D., Ulevitch R.J., Han J.;
RT "Primary structure of BMK1: a new mammalian map kinase.";
RL Biochem. Biophys. Res. Commun. 213:715-724(1995).
CC -!- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND
NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.
CC -!- ENZYME REGULATION: Activated by tyrosine and threonine
phosphorylation (By similarity).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN
HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE
IN LIVER.
CC -!- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN
TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.
CC -!- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN
THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY
ROLE, IS ABSENT.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U25278; AAA81381.1; -
EMBL; U29725; AAA82931.1; -
EMBL; U29726; AAA82932.1; -
EMBL; U29727; AAA82933.1; -
PIR; B56708; B56708.
HSSP; P24941; 1HCL.
Genew; HGNC:6880; MAPK7.
MIM; 602521; -
GO; GO:0004707; F:MAP kinase activity; TAS.
GO; GO:0007165; P:signal transduction; TAS.
InterPro; IPR003527; MAP_kin.

DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; prot_kinase; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS01351; MAPK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Cell cycle;
KW Phosphorylation.
FT DOMAIN 54 346 PROTEIN KINASE.
FT DOMAIN 337 340 POLY-ALA.
FT DOMAIN 433 464 PRO-RICH 1.
FT DOMAIN 520 523 POLY-ARG.
FT DOMAIN 577 699 PRO-RICH 2.
FT NP_BIND 60 68 ATP (BY SIMILARITY).
FT BINDING 83 83 ATP (BY SIMILARITY).
FT ACT_SITE 181 181 BY SIMILARITY.
FT MOD_RES 218 218 PHOSPHORYLATION (ACTIVATES THE KINASE)
(BY SIMILARITY).
FT MOD_RES 220 220 PHOSPHORYLATION (ACTIVATES THE KINASE)
(BY SIMILARITY).
FT CONFLICT 19 32 AREGTRPRHRLCS -> GPVKVEPHTAASVA (IN
REF. 2).
FT CONFLICT 609 609 L -> V (IN REF. 2).
SQ SEQUENCE 815 AA; 88636 MW; 379AD69803207CCF CRC64;

Query Match 15.7%; Score 161.5; DB 1; Length 815;
Best Local Similarity 32.1%; Pred. No. 0.069;
Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;
QY 17 RSTAPLNTQISALRNDPKPLP-----QQPPAPANQDNSSQNTRLQPTPIP 63
Db 576 RPAAPALTSVPAPAPAPTPTPTPVQPTSPPPGPLAQPTGPQPSAGSTSGVPVQPACPPP 635
QY 64 APAPKPAAPRPLDRESPG-----VENKLIPSVGSPASSTP-----LPPDGT 105
Db 636 GPAPHTGPPGPIPVAPPQIATSTLLAAQSLVPPPPGLPGSSTGVLGYFFPGLPPPPDA 695
QY 106 G--PNS----TPNNRAVTPVSQGSNSSADPKAPPPPPVSSGEPPTLG 147
Db 696 GGAPQSSMSESPDNLVT--QQLSKSQVEDPL----PPVFSGTPKSG 737

RESULT 4
GPI_CHLRE
ID GPI_CHLRE STANDARD; PRT; 555 AA.
AC O9FPQ6; Q03927;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpl precursor (Hydroxyproline-rich
glycoprotein 1).
GN Gpl.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21159092; PubMed=11258910;
RA Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RA Goodenough U.W.;
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
plant hydroxyproline-rich glycoproteins.";
RL Biochemistry 40:2978-2987(2001).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.E.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
encoding cell wall hydroxyproline-rich glycoproteins.";

```

Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
-!- FUNCTION: Major component of the outer cell wall W6 (crystalline)
layer.
-!- SUBUNIT: Associates with GP2 and GP3.
-!- PTM: N-glycosylated and O-glycosylated.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF309494; AAG45420.1; -.
EMBL; M58496; AAA69706.1; ALT_SEQ.
GlycoSuiteDB; Q9FPQ6; -.
InterPro; IPR002965; P_rich_extensin.
InterPro; IPR003882; Pistil_extensin.
PRINTS; PR01217; PRICHEXTENSIN.
PRINTS; PR01218; PSTLEXTENSIN.
Glycoprotein; Repeat; Signal.
SIGNAL 1 29 POTENTIAL.
CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GP1.
DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSPX REPEATS.
DOMAIN 259 279 POLY-PRO.
CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;

Query Match 15.5%; Score 159.5; DB 1; Length 555;
Best Local Similarity 32.8%; Pred. No. 0.063;
Matches 38; Conservative 6; Mismatches 49; Indels 23; Gaps 3;

QY 33 PKPLPQQPPAPANQDNSSQNTRLQPTPIIP-----APAPKPAAPRPLDRESPGVENKLI 88
Db 179 PSPSPVPVPPSPAPP-----SPAPVPVPPSPAPPSPAPPVPPSPAPPSPAPPSP 228

QY 89 PSVGSPASSTPLPPDGTGNSTPNRAVTPVSGNSSSADPKAPPPPVSSGEPP 144
Db 229 PSPAPPSPPSPAPPSPVPVPPSPAPPSPA-----PPSPKPPAPPVPPSPPPPPPP 275

RESULT 5
ENAH_MOUSE STANDARD; PRT; 802 AA.
AC Q03173; P70430; P70431; P70432; P70433;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Enabled protein homolog (NPC derived proline-rich protein 1) (NDPP-1).
GN ENAH OR MENA OR NDPF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=93041923; PubMed=1420303;
RA Sazuka T., Tomooka Y., Kathju S., Ikawa Y., Noda M., Kumar S.;
RT "Identification of a developmentally regulated gene in the mouse
central nervous system which encodes a novel proline rich protein.";
RL Biochim. Biophys. Acta 1132:240-248(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4 AND 5), FUNCTION, AND SUBCELLULAR
LOCATION.
RC TISSUE=Brain;
RX MEDLINE=97015079; PubMed=8861907;
RA Gertler F.B., Niebuhr K., Reinhard M., Wehland J., Soriano P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the
control of microfilament dynamics.";
```

```

Cell 87:227-239(1996).
[3]
FUNCTION, AND SUBUNIT.
MEDLINE=99166867; PubMed=10069337;
Lanier L.M., Gates M.A., Witke W., Menzies A.S., Wehman A.M.,
Macklis J.D., Kwiatkowski D., Soriano P., Gertler F.B.;
"Mena is required for neurulation and commissure formation.";
Neuron 22:313-325(1999).
-!- FUNCTION: May be involved in microfilament assembly and cell
motility. Induces the formation of F-actin rich outgrowths in
fibroblasts. Required for neurulation and commissure formation.
-!- SUBUNIT: Binds profilin.
-!- SUBCELLULAR LOCATION: Localized to focal adhesions and, to a
lesser extent, leading edges and stress fibers.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=5;
Name=5; Synonyms=Mena+++;
IsoId=Q03173-1; Sequence=Displayed;
Name=1;
IsoId=Q03173-2; Sequence=VSP_007255, VSP_007256;
Name=2; Synonyms=Mena;
IsoId=Q03173-3; Sequence=VSP_007259, VSP_007260;
Name=3; Synonyms=Mena+;
IsoId=Q03173-4; Sequence=VSP_007259;
Name=4; Synonyms=Mena++;
IsoId=Q03173-5; Sequence=VSP_007257, VSP_007258;
-!- TISSUE SPECIFICITY: IN THE HEART AND TESTIS AND LESS SO IN THE
LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS AND BRAIN.
-!- SIMILARITY: Contains 1 WH1 domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D10727; BAA01570.1; -.
EMBL; U72520; AAC52863.1; -.
EMBL; U72521; AAC52864.1; -.
EMBL; U72522; AAC52865.1; -.
EMBL; U72523; AAC52866.1; -.
PIR; S27200; S27200.
MGD; MGI:108360; Enah.
InterPro; IPR001960; WH1.
Pfam; PF00568; WH1; 1.
SMART; SM00461; WH1; 1.
Developmental protein; Neurogenesis; Alternative splicing.
DOMAIN 1 108 WH1.
FT DOMAIN 442 464 POLY-PRO.
FT DOMAIN 542 552 POLY-PRO.
FT DOMAIN 562 574 POLY-PRO.
FT DOMAIN 578 589 POLY-PRO.
FT DOMAIN 593 605 POLY-PRO.
FT VARSPLIC 1 412 Missing (in isoform 1).
FT VARSPLIC 535 631 /FTid=VSP_007255.
FT VARSPLIC 117 131 PQGLVGGPPAPPPPPPLPSPGAYASALPPPPGPPPPPLPS
FT TISSUE=Brain; TGPPPPPPPPPLPNQAPPPPPPPAPPLPASGIFSGSTSE
FT MEDLINE=93041923; PubMed=1420303; DNRPLTGLAAIAGA -> HRAISWDHLHLRHPLSHQALP
FT Sazuka T., Tomooka Y., Kathju S., Ikawa Y., Noda M., Kumar S.; TPQHLLPQDPLHHLHCHPLVLLHLLHPLHPLFLIKLLPLP
FT "Identification of a developmentally regulated gene in the mouse HLLPLSPHLEFSLDPROKTIAFNWTCSNCGS (in
FT central nervous system which encodes a novel proline rich protein."; isoform 1).
FT Biochim. Biophys. Acta 1132:240-248(1992). /FTid=VSP_007256.
FT [2] Missing (in isoform 4).
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4 AND 5), FUNCTION, AND SUBCELLULAR /FTid=VSP_007257.
RC LOCATION. CLPC -> VFYL (in isoform 4).
RX TISSUE=Brain; /FTid=VSP_007258.
RA MEDLINE=97015079; PubMed=8861907; Missing (in isoform 2 and isoform 3).
RA Gertler F.B., Niebuhr K., Reinhard M., Wehland J., Soriano P.; /FTid=VSP_007259.
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the Missing (in isoform 2).
RT control of microfilament dynamics."; /FTid=VSP_007260.
SQ SEQUENCE 802 AA; 85844 MW; 592BB975EE20F77F CRC64;
```

Query Match 15.4%; Score 158; DB 1; Length 802;
Best Local Similarity 26.3%; Pred. No. 0.1;
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;

QY 13 NKTERSTAPLNT-----QISALRNDP----- 33
||| :|||
Db 348 NKNSRPSPVNTSSQPPAAKSCAWPTSNFSLPSPIMISSPPGKATGPRPVLPCVS 407
||| :|||
QY 34 KPLPQQPPAPANQDN-----SSQNTRLQPTTPIAPAPKPAAPRPL----- 76
||| :|||
Db 408 SPVPQMPSPPTAPNGSLDSVTYPVSPPTSGFAAPPPPPPPPPPPPPPLPPLPLPLAS 467
||| :|||
QY 77 -----DRESPGVENKLI-----PSVGSPPASS-TPLPPD-GTGNSTPNNRV- 116
||| :|||
Db 468 LSHCGSQASPPPGTPLASTPSSKPSVLPSPSACAPASAEETPLNPELGDSSASEPGLQAAS 527
||| :|||
QY 117 ----TPVSQGSNSSADPKAPPPPPVSSGE-----PPTLGENP 150
||| :|||
Db 528 QPAESPTQG--LVLGPPAPPPPPPLPSGPPAYASALPPPPGPPP 569
||| :|||

RESULT 6
APG_BRANA STANDARD; PRT; 449 AA.
AC P40603;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Anter-specific proline-rich protein APG (Protein CEX) (Fragment).
GN APG OR CEX.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94004980; PubMed=8401599;
RA Roberts M.R., Foster G.D., Blundell R.P., Robinson S.W., Kumar A.,
RA Draper J., Scott R.;
RT "Gametophytic and sporophytic expression of an anther-specific
RT Arabidopsis thaliana gene."
RL Plant J. 3:111-120(1993).
CC -!- TISSUE SPECIFICITY: FOUND IN ANTHER, ONLY IN MALE FERTILE PLANTS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN MALE GAMETOGENESIS, DURING
CC MICROSPORE DEVELOPMENT. HIGHER EXPRESSION IS FOUND DURING
CC MICROSPORE MITOSIS WITH A DRAMATIC DECLINE DURING POLLEN
CC MATURATION.
CC -!- SIMILARITY: BELONGS TO THE "GDSL" FAMILY OF LIPOLYTIC ENZYMES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X60376; CAA42924.1; -.
PIR; S16748; S16748.
DR InterPro; IPR001087; Lipase_GDSL.
DR InterPro; IPR003882; Pistil_extensin.
DR Pfam; PF00657; Lipase_GDSL; 1.
DR PRINTS; PR01218; PSTLEXTENSIN.
DR PROSITE; PS01098; LIPASE_GDSL_SER; 1.
FT NON_TER 1
FT ACT_SITE 132 132 BY SIMILARITY.
FT ACT_SITE 428 428 POTENTIAL.
SQ SEQUENCE 449 AA; 48779 MW; 9EFB6A3AB28EEA15 CRC64;

Query Match 15.2%; Score 156; DB 1; Length 449;
Best Local Similarity 33.3%; Pred. No. 0.079;

Matches 39; Conservative 9; Mismatches 57; Indels 12; Gaps 3;
QY 33 PKLPLPQQPPAPANQDNSSQNTRLQPTTPIAPAPKPAAPRPLDRESPGVENKLIPSVG 92
||| :|||
Db 2 PKPQPKPPPKPQKPPPPAPTSPCPPQPKPQKPPPPAPTSPCPPQPKPQKPPPPAPG 61
||| :|||
QY 93 -----SPASSTPLPPDGTGPNSTNNRAVTPVSGSNSSSADPKAPPPPPVSSGEPP 144
||| :|||
Db 62 PSPKPGSPSPPKPPSPAPKVP-----PPSPKPSPPKPPAPSPKP-SPKPP 111
||| :|||

RESULT 7
WAIP_HUMAN STANDARD; PRT; 503 AA.
ID WAIP_HUMAN
AC O43516; Q15220; Q9UNP1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Wiskott-Aldrich syndrome protein interacting protein (WASP interacting
DE protein) (PRPL-2 protein).
GN WASPIP OR WIP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98070810; PubMed=9405671;
RA Ramesh N., Anton I.M., Hartwig J.H., Geha R.S.;
RT "WIP, a protein associated with Wiskott-Aldrich syndrome protein,
RT induces actin polymerization and redistribution in lymphoid cells."
RL Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Tonsil;
RA Kreideweiss S., Delany-Heiken P., Nordheim A., Ruhlmann A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.
RX MEDLINE=99218549; PubMed=10202051;
RA Stewart D.M., Tian L., Nelson D.L.;
RT "Mutations that cause the Wiskott-Aldrich syndrome impair the
RT interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP
RT interacting protein."
RL J. Immunol. 162:5019-5024(1999).
CC -!- FUNCTION: MAY HAVE DIRECT ACTIVITY ON THE ACTIN CYTOSKELETON.
CC INDUCES ACTIN POLYMERIZATION AND REDISTRIBUTION.
CC -!- SUBUNIT: BINDS TO WASP, PROFILIN AND ACTIN.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PERIPHERAL BLOOD
CC MONONUCLEAR CELLS, SPLEEN, PLACENTA, SMALL INTESTIN, COLON,
CC THYMUS. LOWER EXPRESSION IN OVARY, HEART, BRAIN, LUNG, LIVER,
CC SKELETAL MUSCLE, KIDNEY, PANCREAS, PROSTATE, TESTIS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF031588; AAC03767.1; -.
EMBL; AF106062; AAD45972.1; -.
EMBL; X86019; CAA60014.1; -.
DR Genew; HGNC:12736; WASPIP.
DR MIM; 602357; -.
DR GO; GO:0015629; C:actin cytoskeleton; TAS.
DR GO; GO:0003779; F:actin binding activity; TAS.
DR GO; GO:0005522; F:profilin binding activity; TAS.
DR GO; GO:0008154; P:actin polymerization and/or depolymerization; TAS.
DR GO; GO:0006461; P:protein complex assembly; TAS.
DR InterPro; IPR003124; WH2.
DR Pfam; PF02205; WH2; 1.

DR SMART; SM00246; WH2; 1.
KW Actin-binding; Repeat; polymorphism.
FT DOMAIN 2 13 POLY-PRO.
FT DOMAIN 64 96 GLY-RICH.
FT DOMAIN 241 244 POLY-SER.
FT DOMAIN 264 433 PRO-RICH.
FT REPEAT 352 361 XRSXPXPXP MOTIF 1.
FT REPEAT 374 383 XRSXPXPXP MOTIF 2.
FT REPEAT 410 419 XRSXPXPXP MOTIF 3.
FT SITE 45 48 ACTIN BINDING.
FT VARIANT 495 495 G -> A.
FT /FTID=VAR_010295.
FT CONFLICT 303 309 PHRPHLR -> SSQAPP (IN REF. 3).
FT CONFLICT 360 360 P -> PV (IN REF. 2).
FT CONFLICT 487 503 GSNRRRGGLPPPIPR -> EYFCQGF (IN REF. 2).
SQ SEQUENCE 503 AA; 51489 MW; 43EB88674DD3BF1A CRC64;

Query Match 14.7%; Score 151.5; DB 1; Length 503;
Best Local Similarity 26.5%; Pred. No. 0.15;
Matches 60; Conservative 25; Mismatches 84; Indels 57; Gaps 11;

QY 12 NNK-----TERSTAPLNTQISALRNDKPLPQPPAPANQDQNS-----SQNTRL-QPT 59
Db 289 NNKPPVPSTPRPSAPHRPH---LRPPPPSRPGPPPLPPSSSGNDETPLRLQRLNLSST 345
QY 60 PPIAPAPKPAAPRRPLDRSPGVENKLIPIVSGSPASSTPLPPDGTGPNSTPNNRAV--- 116
Db 346 PPLSPGSRGSLPPPPSERPPPPVRD-----PPGRSGPLPPPPVSRNGSTSRALPAT 398
QY 117 --TPVSGSNSSADPK-----APPPPVSSGEPPTLGENPDGLSQEQLEHR-- 161
Db 399 PQLPSRSGVDSRSGRPPPLPPDRPSAGAPPPPPSTSIIRNGFQDSP---CEDEWESREY 455
QY 162 -----ERSLQTLRDIQRLMFLPDEKEFTCA---QSGGPPQNP 194
Db 456 FHPISDLPPPEYVQTTKSYPSKLARNESR-SGSNRRRGGGPPPLPP 500

RESULT 8
PRP2_HUMAN
ID PRP2_HUMAN STANDARD; PRT; 251 AA.
AC P02812;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein precursor (Clone CP7) [Contains: Basic
peptide P-F] (Fragment).
GN PRB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85289325; PubMed=29933301;
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT "Differential RNA splicing and post-translational cleavages in the
human salivary proline-rich protein gene system.";
RL J. Biol. Chem. 260:11123-11130(1985).
RN [2]
RP SEQUENCE OF 134-194.
RX MEDLINE=83265674; PubMed=6874669;
RA Saitoh E., Isemura S., Sanada K.;
RT "Complete amino acid sequence of a basic proline-rich peptide, P-F,
from human parotid saliva."
RL J. Biochem. 93:883-888(1983).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

CC EMBL; K03208; AAA60189.1; -
DR PIR; E25372; PIHUPF.
DR Genew; HGNC:9338; PRB2.
DR MIM; 168810; -
KW Repeat; Parotid gland; Multigene family.
FT NON_TER 1
FT CHAIN 134 194 BASIC PEPTIDE P-F.
FT REPEAT <1 9
FT REPEAT 10 71
FT REPEAT 72 133
FT REPEAT 134 195
SQ SEQUENCE 251 AA; 24641 MW; D779F590C0EBF30B CRC64;

Query Match 14.6%; Score 150.5; DB 1; Length 251;
Best Local Similarity 31.2%; Pred. No. 0.091;
Matches 43; Conservative 9; Mismatches 61; Indels 25; Gaps 5;

QY 31 NDPK--PLPQPPAPANQDQNSQNTLQPTTP-----IPAPAPKPAAPRP 75
Db 45 NKPGPPPPGKPGQPPPGDNKSQSARSPPGKPGQPPPGGNGNQPGPPPPGKPGQPPPPQ 104
QY 76 LDRESPGVENKLIPIVSGSPASSTPLPPDGTGPNSTPNNRAVTPVSGSNSSADPKAPP 135
Db 105 GDNKSQGP-----PPGKPGQPPPGGSKSRSSRSPGKPGQPPPPQGGN-----QPQGGPP 155
QY 136 PP-VSSGEPPTLGENPDG 152
Db 156 PPGKPGQPPPGGNGKPKQ 173

RESULT 9
RXRB_HUMAN
ID RXRB_HUMAN STANDARD; PRT; 533 AA.
AC P28702; P28703;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Retinoic acid receptor RXR-beta.
GN RXRB OR NR2B2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RX MEDLINE=92253386; PubMed=1315958;
RA Fleischhauer K., Park J.H., Disanto J.P., Marks M.S., Ozato K.,
Yang S.Y.;
RT "Isolation of a full-length cDNA clone encoding a N-terminally
variant form of the human retinoid X receptor beta."
RL Nucleic Acids Res. 20:1801-1801(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=92127595; PubMed=1310259;
RA Leid M., Kastner P., Lyons R., Nakshatri H., Saunders M.,
Zacharewski T., Chen J.Y., Staub A., Garnier J.M., Mader S.,
Chambon P.;
RT "Purification, cloning, and RXR identity of the HeLa cell factor with
which RXR or TR heterodimerizes to bind target sequences
efficiently."
RL Cell 68:377-395(1992).
RN [3]
RP REVISIONS.
RX MEDLINE=93046692; PubMed=1330328;
RA Leid M., Kastner P., Lyons R., Nakshatri H., Saunders M.,
Zacharewski T., Chen J.Y., Staub A., Garnier J.M., Mader S.,
Chambon P.;
RL Cell 71:887-887(1992).
RN [4]
RP SEQUENCE FROM N.A.

RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
RL multidomain proteins present in human and rodent brain.";
RN J. Biol. Chem. 274:32997-33001(1999).
RN [6]
RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
RX MEDLINE=99360651; PubMed=10433269;
RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,
RA Doan A., Aakalu V.K., Lanahan A.A., Sheng M., Worley P.F.;
RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
RT postsynaptic density proteins.";
RL Neuron 23:583-592(1999).
RN [7]
RP INTERACTION WITH SPTAN1.
RX MEDLINE=21523912; PubMed=11509555;
RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,
RA Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;
RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the
RT multidomain Shank protein family interact with the cytoskeletal
RT protein alpha-fodrin.";
RL J. Biol. Chem. 276:40104-40112(2001).
RN [8]
RP FUNCTION.
RX MEDLINE=21389514; PubMed=11498055;
RA Sala C., Piech V., Wilson N.R., Passafaro M., Liu G., Sheng M.;
RA "Regulation of dendritic spine morphology and synaptic function by
RT Shank and Homer.";
RL Neuron 31:115-130(2001).
RN [9]
RP REVIEW.
RX MEDLINE=20267867; PubMed=10806096;
RA Sheng M., Kim E.;
RA "The Shank family of scaffold proteins.";
RL J. Cell Sci. 113:1851-1856(2000).
CC -!- FUNCTION: Seems to be an adapter protein in the postsynaptic
CC density (PSD) of excitatory synapses that interconnects receptors
CC of the postsynaptic membrane including NMDA-type and metabotropic
CC glutamate receptors, and the actin-based cytoskeleton. May play a
CC role in the structural and functional organization of the
CC dendritic spine and synaptic junction. Overexpression promotes
CC maturation of dendritic spines and the enlargement of spine heads
CC via its ability to recruit Homer to postsynaptic sites, and
CC enhances presynaptic function.
CC -!- SUBUNIT: May homomultimerize via its SAM domain. Interacts with
CC SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with
CC DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via
CC the PDZ domain (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
CC neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.
CC -!- ALTERNATIVE PRODUCTS;
CC Event=Alternative splicing; Named isoforms=5;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9WV48-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9WV48-2; Sequence=VSP_006072, VSP_006073;
CC Name=3;
CC IsoId=Q9WV48-3; Sequence=VSP_006074;
CC Name=4; Synonyms=A;
CC IsoId=Q9WV48-4; Sequence=VSP_006075;
CC Name=5;
CC IsoId=Q9WV48-5; Sequence=VSP_006076, VSP_006077;
CC -!- TISSUE SPECIFICITY: Expressed only in brain (neuropil of cortex,
CC CA1 region hippocampus and molecular layer of cerebellum).
CC -!- DEVELOPMENTAL STAGE: Expression increases from low levels at birth
CC to high levels at 3-4 weeks before dropping slightly in adulthood.
CC Expressed in the cortex and the molecular layer of the cerebellum
CC at postnatal day 7. Isoform 2 expression does not change during
CC development of both cortex and cerebellum. Isoform 4 expression
CC decreases significantly during development of cortex but not
CC cerebellum.
CC -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.

CC CC
CC -!- SIMILARITY: Contains 7 ANK repeats.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF102855; AAD04569.2; -
DR EMBL; AF131951; AAD29417.1; ALT_INIT.
DR EMBL; AF159046; AAD42975.1; -
DR EMBL; AF141904; AAF02498.1; ALT_INIT.
DR HSSP; P00519; IABL.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; ank; 7.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00536; SAM; 1.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50106; PDZ; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
KW ANK repeat; SH3 domain; Repeat; Alternative splicing.
FT REPEAT 195 210 ANK 1.
FT REPEAT 212 245 ANK 2.
FT REPEAT 246 278 ANK 3.
FT REPEAT 279 312 ANK 4.
FT REPEAT 313 345 ANK 5.
FT REPEAT 346 378 ANK 6.
FT REPEAT 379 395 ANK 7.
FT DOMAIN 554 613 SH3.
FT DOMAIN 663 757 PDZ.
FT DOMAIN 2104 2167 SAM.
FT DOMAIN 929 932 POLY-PRO.
FT DOMAIN 1010 1015 POLY-HIS.
FT DOMAIN 1022 1027 POLY-HIS.
FT DOMAIN 1194 1199 POLY-GLY.
FT DOMAIN 1850 1860 POLY-PRO.
FT VARSPLIC 1 613 Missing (in isoform 2).
FT VARSPLIC 615 654 /FTId=VSP_006072.
FT FT SQEGRQSRSDKAKRLFRHYTVGSYDSFSDAPSLIDGDSG
FT FT -> MALSAVGGPGGGALPQPPALSSSWPALGPRRRSVWY
FT FT IY (in isoform 2).
FT FT /FTId=VSP_006073.
FT FT Missing (in isoform 3).
FT FT /FTId=VSP_006074.
FT FT Missing (in isoform 4).
FT FT /FTId=VSP_006075.
FT FT LSEDSQTSLLSKPS -> QYRIIVKSSDFGDF (in
FT FT isoform 5).
FT FT /FTId=VSP_006076.
FT FT Missing (in isoform 5).
FT FT /FTId=VSP_006077.
FT FT S -> T (IN REF. 1).
FT FT S -> N (IN REF. 2).
FT FT R -> K (IN REF. 1).
FT FT A -> T (IN REF. 1).
FT FT S -> D (IN REF. 1).
FT FT S -> N (IN REF. 2).

RT HMNA, of chicken and marked decrease of its immunoreactivity during
the progression of S phase.";
J. Cell Sci. 110:3031-3041(1997).
DR EMBL; D88440; BAA24137.1; -
DR InterPro; IPR002965; P_rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
FT NON_TER 1 1151
FT SEQUENCE 1151 AA; 109708 MW; 2403F6835F9A2AB3 CRC64;
SQ
Query Match 15.8%; Score 162.5; DB 13; Length 1151;
Best Local Similarity 33.1%; Pred. No. 0.00052;
Matches 45; Conservative 12; Mismatches 40; Indels 39; Gaps 6;
QY 33 PKPLPQPPAPANQDNSSQNTLRQLPTP-PIAPAPKPAAPRP-----LDRESPGVE 84
Db 3 PTPPPRNPPTP-----PPAPSPAPAPAPAPAPAPPRPKWVPIAELHPAAPP 49
QY 85 NKLIPSVGSPASSTPLPPDGTGPNSTPNNAKAVTPVSQGSNSSSADPKAP---PPPPVSSG 141
Db 50 PKWVPIGGAP-----PPPGTEP-----TPPSKPTDGADAAPKASAEALSPPPASP 94
QY 142 EPPTLGENPDGLSQEQ 157
Db 95 SPDPGPKAPSGAGEAE 110
RESULT 8
Q96G51 PRELIMINARY; PRT; 816 AA.
AC Q96G51; Q96991;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Mitogen-activated protein kinase 7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas, and Muscle;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; BC009963; AAH09963.1; -
DR EMBL; BC007992; AAH07992.1; -
DR EMBL; BC007404; AAH07404.1; -
DR EMBL; BC030134; AAH30134.1; -
DR HSSP; P24941; 1BUH.
DR InterPro; IPR003527; MAP_kin.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRODOM; PD000001; Prot_kinase; 1.
DR PROSITE; PS01351; MAPK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 816 AA; 88386 MW; 27729FE31658CE3B CRC64;

Query Match 15.7%; Score 161.5; DB 4; Length 816;
Best Local Similarity 32.1%; Pred. No. 0.00044;
Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;

QY 17 RSTAPLNTQISALRNDPKLP-----QPPAPANQDNSSQNTLRQLPTPIP 63
Db 577 RPAAPALTSVPAPAPAPTPTPTVQPTSPPPGPPVQAQTPGQSGSTSGPVPPACPPP 636
QY 64 APAPKPAAPRPLDRESPG-----VENKLIPSVGSPASSTP-----LPPDGT 105
Db 637 GPAPHTGPPGPIPVAPPQIATSTSLAAQSLVPPGGLPGSSTPGVLPYFPFGLPPDA 696
QY 106 G--PNS---TPNNRAVTPVSQGSNSSSADPKAPPPPPVSSGEPPTLG 147
Db 697 GGAPQSSMSESPDNLVT--QQLSKSQVEDPL-----PPVFSGTGPKSG 738
RESULT 9
Q9DGT6 PRELIMINARY; PRT; 2321 AA.
AC Q9DGT6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE RS1 immediate-early gene transactivator ICP4-like protein (RS1
DE immediate-early gene transactivator-like protein).
GN MDV084 OR MDV100.
OS Turkey herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCBI_TaxID=10390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Md5;
RX MEDLINE=20392152; PubMed=10933706;
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
RT "The genome of a very virulent Marek's disease virus.";
RL J. Virol. 74:7980-7988(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Md5;
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF243438; AAG14284.1; -
DR EMBL; AF243438; AAG14273.1; -
DR InterPro; IPR005205; Herpes_ICP4_C.
DR InterPro; IPR005206; Herpes_ICP4_N.
DR InterPro; IPR002965; P_rich_extensn.
DR Pfam; PF03585; Herpes_ICP4_C; 1.
DR Pfam; PF03584; Herpes_ICP4_N; 1.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 2321 AA; 252959 MW; EF130EB0C4503003 CRC64;

Query Match 15.6%; Score 160.5; DB 12; Length 2321;
Best Local Similarity 32.4%; Pred. No. 0.0015;
Matches 60; Conservative 21; Mismatches 53; Indels 51; Gaps 13;

QY 12 NNKTERSTAPLNTQISALRNDPKLP-----PQOPP-APANQD-----QNSSQNTLR- 56
Db 241 NRASERQLA--DTAASALR-APSPVFWSAFDSRYPHILAPANQSNSDPLCPETSTASAQIL 297
QY 57 ---QPTPPI---PAPAPKPAAP----PRPLDRESPGVENKLIPSVGSPAS-----ST 98
Db 298 HTNSPTPTSTSPAPISPTQPPACLPSPAPISPPVQPPALLPLIFSPITPVEFIQPPSP 357
QY 99 PLPPDGTGP-----NSTPNNAVTPVSQGSNSS-----SADPKAPP---PPPVSSGEPPTL 146
Db 358 PSPPQAPSPPAHSSSSCSPLAPSLSSPLSSPLSPQLSAPVSPSPSPPLSPGE---L 414
QY 147 GENPD 151
Db 415 APSPD 419

RESULT 10
Q8DGU1 PRELIMINARY; PRT; 488 AA.
ID Q8DGU1

Query Match 15.4%; Score 158.5; DB 10; Length 1188;
Best Local Similarity 32.0%; Pred. No. 0.0011;
Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;

QY 14 KTERSTAPLNTQISALRNDPKPLP--QQPPAPANQDNSSQNTRLQP---TPPIPAPAPK 68
Db 961 KSSPPAPVNLPPPEVKSSPPPTPVSSPPAPKSPPPAPMSSPPPEVKSPPPAPVSS 1020

QY 69 P-----AAPRPLDRESQVENKLIIPSVGSPASSTPLP---PDGTGPNSTPNNRATPV 119
Db 1021 PPPPVKSPPPAPVSSPPPVKS---PPPPAPVSSPPPPVKSPPPAPISPPPPVKSP 1077

QY 120 SQGSNSSSADP-KAPPPP-pvssGEPPTLGENP 150
Db 1078 PPAPVSSPPPPVKSPPPAPVSSPPPIKSPPP 1110

RESULT 15
P70433
ID P70433 PRELIMINARY; PRT; 802 AA.
AC P70433; P70430; P70431; P70432;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE ENABLED homolog.
GN ENAH OR MENA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=BRAIN;
RX MEDLINE=97015079; PubMed=8861907;
RA Gertler F.B., Niebuhr K., Reinhard M., Wehland J., Soriano P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the
RT control of microfilament dynamics.";
RL Cell 87:227-239(1996).
CC -!- FUNCTION: MAY BE INVOLVED IN MICROFILAMENT ASSEMBLY AND CELL
CC MOTILITY. INDUCES THE FORMATION OF F-ACTIN RICH OUTGROWTHS IN
CC FIBROBLASTS. BINDS PROFILIN.
CC -!- SUBCELLULAR LOCATION: LOCALIZED TO FOCAL ADHESIONS AND, TO A
CC LESSER EXTENT, LEADING EDGES AND STRESS FIBERS.
CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; MENA, MENA+, MENA++ AND MENA+++
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; U72520; AAC52863.1; -.
DR EMBL; U72521; AAC52864.1; -.
DR EMBL; U72522; AAC52865.1; -.
DR EMBL; U72523; AAC52866.1; -.
DR MGD; MGI:108360; Enah.
DR InterPro; IPR000697; EVH1.
DR InterPro; IPR000156; Ran_BP1.
DR InterPro; IPR001960; WH1.
DR Pfam; PF00568; WH1; 1.
DR SMART; SM00160; RanBD; 1.
DR SMART; SM00461; WH1; 1.
KW Alternative splicing; Phosphorylation.
FT DOMAIN 442 464 POLY-PRO.
FT DOMAIN 542 552 POLY-PRO.
FT DOMAIN 562 574 POLY-PRO.
FT DOMAIN 578 589 POLY-PRO.
FT DOMAIN 593 605 POLY-PRO.
FT DOMAIN 686 689 POLY-ARG.
FT MOD_RES 255 255 PHOSPHORYLATION (BY CAPK AND CGPK) (BY
FT SIMILARITY).
FT MOD_RES 637 637 PHOSPHORYLATION (BY CAPK AND CGPK) (BY
FT SIMILARITY).
FT VARSPLIC 117 131 MISSING (IN ISOFORM MENA++).
FT VARSPLIC 117 135 MISSING (IN ISOFORM MENA AND ISOFORM
FT MENA+).
FT VARSPLIC 132 135 C1FC -> VFYL (IN ISOFORM MENA++).
FT VARSPLIC 259 500 MISSING (IN ISOFORM MENA).
SQ SEQUENCE 802 AA; 85844 MW; 592BB975EE20F77F CRC64;

Query Match 15.4%; Score 158; DB 11; Length 802;
Best Local Similarity 26.3%; Pred. No. 0.00079;
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;

QY 13 NKTERSTAPLNT-----QISALRNDP----- 33
Db 348 NKNSRPSSPVNTSSQPPAAKSCAWPTSNFSLPPSPIMISSPPGKATGPRPVLPVCVS 407

QY 34 KPLPQQPPAPANQDN-----SSQNTRLQPTPIAPAPKPAAPRPL----- 76
Db 408 SPVFQMPSPPTAPNGSLDSVTYPVSPPTSGPAAPPPPPPPPPPPPPPPPLPLPLPLAS 467

QY 77 -----DRESPGVENKLI-----PSVGSPASS-TPLPPD-GTGNSTPNNRV- 116
Db 468 LSHCGSQASPPPGTPLASTPSSKPSVLPSPSAGAPASAEPTLNPELGDSSASEPGLQAAS 527

QY 117 ----TPVSQGSNSSSADPKAPPPPPVSSGE-----PPTLGENP 150
Db 528 QPAESPTPG--LVLGPPAPPPPPPPPLPSGPPAYASALPPPPGPPP 569

Search completed: October 15, 2003, 10:35:50
Job time : 115.475 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 21.5175 Seconds
(without alignments)
258.182 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383
Perfect score: 178
Sequence: 1 DGLSQEQLHRRSLQTLRDQLRMLFPDEKEFTGA 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				A_Geneseq_19Jun03:*
1:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1980.DAT:*		
2:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1981.DAT:*		
3:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1982.DAT:*		
4:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1983.DAT:*		
5:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1984.DAT:*		
6:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1985.DAT:*		
7:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1986.DAT:*		
8:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1987.DAT:*		
9:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1988.DAT:*		
10:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1989.DAT:*		
11:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1990.DAT:*		
12:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1991.DAT:*		
13:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1992.DAT:*		
14:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1993.DAT:*		
15:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1994.DAT:*		
16:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1995.DAT:*		
17:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1996.DAT:*		
18:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1997.DAT:*		
19:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1998.DAT:*		
20:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA1999.DAT:*		
21:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA2000.DAT:*		
22:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA2001.DAT:*		
23:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA2002.DAT:*		
24:	/SIDS1/gcgdata/	geneseq/geneseq-emb1/AA2003.DAT:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	178	100.0	1426	23	Human legless homo
2	178	100.0	1435	22	Human BCL9 homolog
3	109	61.2	320	23	Mouse beta-catenin
4	109	61.2	1494	23	Mouse beta-catenin
5	107	60.1	738	23	Human beta-catenin
6	107	60.1	1115	23	Human legless homo
7	66	37.1	1429	22	Drosophila melanog
8	66	37.1	1464	23	D. melanogaster lg
9	57.5	32.3	603	24	Aspergillus fumiga

10	57.5	32.3	618	24	ABJ26453	Aspergillus fumiga
11	56	31.5	1294	22	ABB63502	Drosophila melanog
12	55	30.9	294	22	AAB95073	Human protein sequ
13	55	30.9	390	23	ABG96285	Human ovarian canc
14	55	30.9	433	22	AAG63851	Amino acid sequenc
15	55	30.9	433	22	AAG63852	Amino acid sequenc
16	55	30.9	433	22	AAB68522	Human GTP-binding
17	55	30.9	464	22	AAB92828	Human protein sequ
18	55	30.9	464	23	ABG96281	Human ovarian canc
19	55	30.9	643	23	ABG96282	Human ovarian canc
20	55	30.9	718	23	ABG96284	Human ovarian canc
21	55	30.9	751	23	ABG96283	Human ovarian canc
22	54.5	30.6	174	22	AAB86346	Murine bHLH transcr
23	54.5	30.6	174	22	AAB86348	Murine bHLH transcr
24	53	29.8	819	22	AAU69744	Thermus thermophil
25	52.5	29.5	181	22	AAB86345	Human bHLH transcr
26	52.5	29.5	181	22	AAB86347	Human bHLH transcr
27	52	29.2	959	21	AAV53051	Human secreted pro
28	51.5	28.9	210	23	ABP43969	Adenomatous polypo
29	51	28.7	329	21	AAG13171	Arabidopsis thalia
30	51	28.7	434	22	AAU33491	Enterococcus faeca
31	51	28.7	440	23	ABP47770	Protein #16 relate
32	51	28.7	448	22	AAU35058	Enterococcus faeca
33	51	28.7	500	22	AAG89346	Human secreted pro
34	51	28.7	1464	22	ABB71111	Drosophila melanog
35	51	28.7	1755	20	AAV41139	Mouse mammary tumo
36	50.5	28.4	96	22	ABB69662	Drosophila melanog
37	50.5	28.4	479	23	ABP43965	Unidentified prote
38	50.5	28.4	675	21	AAV54052	An angiogenesis-as
39	50.5	28.4	675	21	AAV54053	A variant of an an
40	50.5	28.4	1183	22	ABB58769	Drosophila melanog
41	50	28.1	187	21	AAB38555	Human secreted pro
42	50	28.1	221	21	AAB07850	Amino acid sequenc
43	50	28.1	223	22	AAB62225	Glycine max glutat
44	50	28.1	275	21	AAB24369	Bacillus subtilis
45	50	28.1	342	20	AAV42781	Human neuronal imm

ALIGNMENTS

RESULT 1

AAAB71229	ID	AAAB71229	standard; Protein; 1426 AA.
XX	XX		
AAAB71229;	AC	AAAB71229;	
XX	XX		
18-NOV-2002 (first entry)	DT		
XX	XX		
Human legless homologue lgs/bcl9 protein.	DE		
XX	XX		
Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;	KW		
tissue proliferation; tumour; cytosolic; cellular disorder; colon;	KW		
blood disorder; cancer; breast; head and neck cancer; brain; thyroid;	KW		
medulloblastoma; skin cancer; tissue regeneration; tissue repair.	KW		
XX	XX		
Homo sapiens.	OS		
XX	XX		
US2002086986-A1.	PN		
XX	XX		
04-JUL-2002.	PD		
XX	XX		
27-JUL-2001; 2001US-0915543.	PF		
XX	XX		
28-JUL-2000; 2000US-221502P.	PR		
XX	XX		
(BASL/) BASLER K.	PA		
(BRUN/) BRUNNER E.	PA		
(FROE/) FROESCH B.	PA		
(KRAM/) KRAMPS T.	PA		
(PETE/) PETER O.	PA		
XX	XX		
Basler K, Brunner E, Froesch B, Kramps T, Peter O;	PI		

XX WPI; 2002-635689/68.
DR N-PSDB; AAF88467.
XX
PT Novel polypeptide useful in therapeutic method for treating disorders
PT of cell fate such as cell differentiation or cell proliferation
XX
PS Example II; Fig 8B; 4lpp; English.
XX
CC This invention describes a novel polypeptide sharing one or more
CC homologous amino acid domains with the legless (lgs) protein, a
CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC involved in the formation and maintenance of spatial arrangements
CC and proliferation of tissues during development, and in the formation
CC and growth of many human tumours. The products of the invention have
CC cytoskeletal activity and can be used to treat cellular disorders, blood
CC disorders and cancers caused by over-stimulation of the Wnt pathway,
CC where the cancerous condition is colon, breast, head and neck, brain,
CC thyroid, medulloblastoma or skin cancer. The product could also be used
CC to promote tissue regeneration and repair. This sequence represents the
CC human legless (lgs) protein homologue lgs/bcl9 described in the
CC disclosure of the invention.
XX
SQ Sequence 1426 AA;
Query Match 100.0%; Score 178; DB 23; Length 1426;
Best Local Similarity 100.0%; Pred. No. 8.6e-16;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGLSQEQLHRRSLQTLRDIQRLFPDEKEFTGA 35
Db 349 DGLSQEQLHRRSLQTLRDIQRLFPDEKEFTGA 383
RESULT 2
ABB11808
ID ABB11808 standard; peptide; 1435 AA.
XX
AC ABB11808;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human BCL9 homologue, SEQ ID NO:2178.
XX
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer.
XX
OS Homo sapiens.
XX
PN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US03800.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX

DR WPI; 2001-457740/49.
DR N-PSDB; ABA09052.
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
PS Claim 20; Page 256-257; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel, many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
XX polypeptide of the invention.
SQ Sequence 1435 AA;
Query Match 100.0%; Score 178; DB 22; Length 1435;
Best Local Similarity 100.0%; Pred. No. 8.7e-16;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGLSQEQLHRRSLQTLRDIQRLFPDEKEFTGA 35
Db 389 DGLSQEQLHRRSLQTLRDIQRLFPDEKEFTGA 423
RESULT 3
AAU78461
ID AAU78461 standard; Protein; 320 AA.
XX
AC AAU78461;
XX
DT 02-JUL-2002 (first entry)
XX
DE Mouse beta-catenin nuclear localised protein #2.
XX
KW Mouse; beta-catenin nuclear localised protein; cancer;
KW gene therapy; EST; expressed sequence tag.
XX
OS Mus musculus.
XX

PD 04-JUL-2002.
XX
PF 27-JUL-2001; 2001US-0915543.
XX
PR 28-JUL-2000; 2000US-221502P.
XX
PA (BASL/) BASLER K.
PA (BRUN/) BRUNNER E.
PA (FROE/) FROESCH B.
PA (KRAM/) KRAMPS T.
PA (PETE/) PETER O.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2002-635689/68.
DR N-PSDB; AAF88466.
XX
PT Novel polypeptide useful in therapeutic method for treating disorders
PT of cell fate such as cell differentiation or cell proliferation -
XX
PS Example II; Fig 2; 41pp; English.
XX
CC This invention describes a novel polypeptide sharing one or more
CC homologous amino acid domains with the legless (lgs) protein, a
CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC involved in the formation and maintenance of spatial arrangements
CC and proliferation of tissues during development, and in the formation
CC and growth of many human tumors. The products of the invention have
CC cytostatic activity and can be used to treat cellular disorders, blood
CC disorders and cancers caused by over-stimulation of the Wnt pathway,
CC where the cancerous condition is colon, breast, head and neck, brain,
CC thyroid, medulloblastoma or skin cancer. The product could also be used
CC to promote tissue regeneration and repair. This sequence represents the
CC Drosophila melanogaster (fruitfly) legless (lgs) protein described in
CC the disclosure of the invention.
XX
SQ Sequence 1464 AA;
Query Match 37.1%; Score 66; DB 23; Length 1464;
Best Local Similarity 31.4%; Pred. No. 3.2;
Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;
QY 1 DGLSQEQLEHRSLSQTLRDIQRMFLFPDEKEFTGA 35
Db 515 ENLTPQQRHREEQLEAKIKKMNQFLFPENENSVGA 549
RESULT 9
ABJ25853
ID ABJ25853 standard; Protein; 603 AA.
XX
AC ABJ25853;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #511.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US13142.
XX
PR 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.

XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer -
XX
PS Disclosure; Page -; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention.
XX
SQ Sequence 603 AA;
Query Match 32.3%; Score 57.5; DB 24; Length 603;
Best Local Similarity 31.4%; Pred. No. 18;
Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1;
QY 1 DGLSQEQLEHRSLSQTLRDIQRMFLFPDEKEFTGA 35
Db 45 DGVETEKIREX---EVEKKLERMLFGDDEGFVGA 76
RESULT 10
ABJ26453
ID ABJ26453 standard; Protein; 618 AA.
XX
AC ABJ26453;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #1111.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US13142.
XX

```

PR 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
XX WPI; 2003-093124/08.
DR
XX
XX New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer -
XX
XX Disclosure; Page -; 175pp; English.
PS
XX
XX The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention.
XX
XX SQ Sequence 618 AA;
Query Match 32.3%; Score 57.5; DB 24; Length 618;
Best Local Similarity 31.4%; Pred. No. 19;
Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1
QY 1 DGLSQQLHRRERSLQTLRDIQRMFLFPDEKEFTGA 35
||: ||: :: : : |||| ||: | ||
Db 45 DGVETEKIREKD--EVEKKLERMLFGDDEGFVGA 76
RESULT 11
ABB63502
ID ABB63502 standard; Protein; 1294 AA.
XX
XX AC ABB63502;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 17298.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN

```


PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 16943; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 294 AA;
 Query Match 30.9%; Score 55; DB 22; Length 294;
 Best Local Similarity 52.2%; Pred. No. 18;
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
 QY 2 GLSQEQLEHRRSLQTLRDIQRM 24
 || || | | :||:|:
 Db 224 GLRTEGLFRRSASVQTVREIQL 246
 RESULT 13
 ABG96285
 ID ABG96285 standard; Protein; 390 AA.
 XX
 AC ABG96285;
 XX
 DT 11-DEC-2002 (first entry)
 XX
 DE Human ovarian cancer marker M355.
 XX
 KW Human; ovarian cancer; marker; cancer; familial history; brain disorder;
 KW central nervous system disorder; bacterial meningitis; viral meningitis;
 KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;
 KW brain herniation; inflammation; encephalitis; testicular disorder;
 KW nontuberculous granulomatous orchitis; connective tissue disorder;
 KW heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;
 KW histological type; carcinogenic; ovarian cancer marker.
 XX
 OS Homo sapiens.
 XX
 PN WO200271928-A2.
 XX
 PD 19-SEP-2002.
 XX
 PF 14-MAR-2002; 2002WO-US07826.
 XX
 PR 14-MAR-2001; 2001US-276025P.
 PR 14-MAR-2001; 2001US-276026P.

PR 10-AUG-2001; 2001US-311732P.
 PR 19-SEP-2001; 2001US-323580P.
 PR 26-SEP-2001; 2001US-324967P.
 PR 26-SEP-2001; 2001US-325102P.
 PR 26-SEP-2001; 2001US-325149P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Monahan JE, Gannavarapu M, Hoersch S, Kamatkar S, Kovattis SG;
 PI Meyers RE, Morrisey MP, Olandt PJ, Sen A, Vieby PO, Mills GB;
 PI Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K;
 XX
 DR WPI; 2002-723277/78.
 DR N-PSDB; ABS76377.
 XX
 PT Assessing whether a patient is afflicted with ovarian cancer, useful in
 PT assessing the stage or progression of the disease, comprises comparing
 PT the expression level of a cancer marker in a sample from a patient and
 PT from a non cancer patient -
 XX
 PS Disclosure; Page 134-135; 481pp; English.
 XX
 CC The present invention relates to a new method for assessing whether a
 CC patient is afflicted with ovarian cancer. The method involves comparing
 CC the expression level of a marker in a patient sample and the normal level
 CC of expression of the marker in a control non-ovarian cancer sample, where
 CC the marker is selected from 363 cancer markers described in the
 CC specification. The method of the invention is useful in diagnosing or
 CC characterising cancer, in detecting the presence of cancer as early as
 CC possible, and the recurrence of ovarian cancer. The method may also be of
 CC particular use with patients having an enhanced risk of developing
 CC ovarian cancer (e.g. patients having a familial history of ovarian
 CC cancer). The cancer markers may be used in the management and treatment
 CC of e.g. brain and central nervous system disorders (e.g. bacterial and
 CC viral meningitis, Alzheimer's disease or Parkinson's disease), brain
 CC disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),
 CC inflammations (e.g. bacterial or viral meningitis or encephalitis),
 CC testicular disorders (e.g. nontuberculous granulomatous orchitis),
 CC connective tissue disorders, or heart disorders (e.g. ischaemic heart
 CC disease or atherosclerosis). The compositions and methods may also be
 CC used in assessing the histological type of neoplasm associated with
 CC ovarian cancer, monitoring the progression of ovarian cancer,
 CC determining whether ovarian cancer has metastasized or is likely to
 CC metastasize, selecting a composition for inhibiting ovarian cancer,
 CC assessing the ovarian carcinogenic potential of a compound, or
 CC inhibiting ovarian cancer or at risk of developing ovarian cancer. The
 CC present amino acid sequence represents one of the ovarian cancer markers
 CC described in the invention.
 XX
 SQ Sequence 390 AA;
 Query Match 30.9%; Score 55; DB 23; Length 390;
 Best Local Similarity 52.2%; Pred. No. 25;
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
 QY 2 GLSQEQLEHRRSLQTLRDIQRM 24
 || || | | :||:|:
 Db 181 GLRTEGLFRRSASVQTVREIQL 203
 RESULT 14
 AAG63851
 ID AAG63851 standard; Protein; 433 AA.
 XX
 AC AAG63851;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Amino acid sequence of human GTPase activating protein GTPAP1.
 XX
 KW GTPase activating protein; GTPAP1; cell signalling; immune disorder;
 KW cell proliferative disorder; cancer; colon cancer; arteriosclerosis;
 KW diabetes; psoriasis; hepatitis; multiple sclerosis; gene therapy.

Search completed: October 15, 2003, 10:30:24
Job time : 23.5175 secs


```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6509155 3068538CD1
US-09-507-765-30
```

Query Match 30.9%; Score 55; DB 4; Length 433;
Best Local Similarity 52.2%; Pred. No. 9.1;
Matches 12; Conservative 4; Mismatches 7; Indels

Qy 2 GLSQEQLHRRSLQTLRDIQRM 24
 || | | | :||:| |:
 Db 224 GLRTEGLFRRSASVOTVREIORL 246

RESULT 3

```

US-09-507-765-31
; Sequence 31, Application US/09507765
; Patent No. 6509155
; GENERAL INFORMATION:
; APPLICANT: Klinger, Tod M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010 US
; CURRENT APPLICATION NUMBER: US/09/507,765
; CURRENT FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6509155 4044
US-09-507-765-31

```

Query Match 30.9%; Score 55; DB 4; Length 433;
Best Local Similarity 52.2%; Pred. No. 9.1;
Matches 12; Conservative 4; Mismatches 7; Indels

QY 2 GLSQEQLHRRERSLQTLRDIQRM 24
 || | | | | : | : | : | : | :
 Db 224 GLRTEGLFRSSASVOTVREIORL 246

RESULT 4

US-09-252-991A-25843
; Sequence 25843, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25843
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25843

Query Match 30.6%; Score 54.5; DB 4; Length 1242;
Best Local Similarity 35.1%; Pred. No. 33;

	Matches	13;	Conservative	9;	Mismatches	6;	Indels	9;	Gaps	1;
Qy	5	QEQLEHRRSL	-----	QTLRDIQRLFPDEKEF	32					
		:		:		:		:		
Db	169	RRQHEQQRSLRSDIAMIEIRSLRDLRLLLFIYOREF	205							

RESULT 5

US-09-651-656-15
; Sequence 15, Application US/09651656
; Patent No. 6340566
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: DETECTION AND QUANTITATION OF SINGLE NUCLEOTIDE
; TITLE OF INVENTION: POLYMORPHISMS, DNA SEQUENCE VARIATIONS, DNA MUTATIONS,
; TITLE OF INVENTION: DNA DAMAGE AND DNA MISMATCHES
; FILE REFERENCE: IL-10689
; CURRENT APPLICATION NUMBER: US/09/651,656
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-651-656-15

Query Match	29.8%;	Score 53;	DB 4;	Length 819;
Best Local Similarity	42.3%;	Pred. No. 34;		
Matches 11;	Conservative	5;	Mismatches 10;	Indels 0;

QY 6 EQLEHRRSLQTLRDIQRMLFPDEKE 31
|:: | :||:| || |:: |
Db 466 EKVPQEXRPVOTLKDPRORYTLPMEKE 491

RESULT 6

```

US-09-650-855-15
; Sequence 15, Application US/09650855
; Patent No. 6365355
; GENERAL INFORMATION:
; APPLICANT: MCCUTHEN-MALONEY, SANDRA
; APPLICANT: LAWRENCE LIVERMORE NATIONAL LABORATORY
; TITLE OF INVENTION: CHIMERIC PROTEINS FOR DETECTION AND QUANTITATION OF DNA
; TITLE OF INVENTION: MUTATIONS, DNA SEQUENCE VARIATIONS, DNA DAMAGE AND DNA
; TITLE OF INVENTION: MISMATCHES
; FILE REFERENCE: IL-10284
; CURRENT APPLICATION NUMBER: US/09/650,855
; CURRENT FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: 60/192,764
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 15
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-650-855-15

```

Query Match 29.8%; Score 53; DB 4; Length 819;
Best Local Similarity 42.3%; Pred. No. 34;
Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0.

QY 6 EQLEHFRSLQTLRDIQRMLFPDEKE 31
|:: | :|||:| || |:: |
Db 466 EKVPQEXRPVOTLKDRORYTLPMEKE 491

RESULT. 7

US-09-252-991A-23906

```
; Sequence 23906, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23906
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23906

Query Match          29.2%; Score 52; DB 4; Length 578;
Best Local Similarity 84.6%; Pred. No. 32;
Matches 11; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY      6 EQLEHRSRSLQTL 18
      | |||||
Db      339 ECLEHRSRSLTL 351

RESULT 8
US-09-252-991A-24834
; Sequence 24834, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24834
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24834

Query Match          28.7%; Score 51; DB 4; Length 589;
Best Local Similarity 37.0%; Pred. No. 44;
Matches 10; Conservative 5; Mismatches 12; Indels 12; Gaps 0;

QY      5 QEQLHRSRSLQTLRDIQRMFLFPDEKE 31
      :| | | | | | | | | | | | | | | |
Db      156 EQQRNHRRLQRYRQAQRGLGFGEDAE 182

RESULT 9
US-09-252-991A-28779
; Sequence 28779, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
```

```
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28779
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28779

Query Match          28.4%; Score 50.5; DB 4; Length 354;
Best Local Similarity 40.0%; Pred. No. 30;
Matches 14; Conservative 6; Mismatches 8; Indels 7; Gaps 2;

QY      2 GLSQE-----QLEHRSRSLQTLRDIQRMFLFPDEKE 31
      ||:| | | | | | | | | | | | | | | |
Db      252 GLAQRLLIPGQAPHRPRTLRAVRDRQR--PDLRQ 284

RESULT 10
US-09-247-373B-54
; Sequence 54, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1108-A
; CURRENT APPLICATION NUMBER: US/09/247,373B
; CURRENT FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 08/924,747
; PRIOR FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
; LENGTH: 221
; TYPE: PRT
; ORGANISM: SOYBEAN
US-09-247-373B-54

Query Match          28.1%; Score 50; DB 3; Length 221;
Best Local Similarity 36.7%; Pred. No. 21;
Matches 11; Conservative 7; Mismatches 8; Indels 4; Gaps 1;

QY      9 EHRERSLQTL----LRDIQRMFLFPDEKEFTG 34
      | | | | | | | | | | | | | | | |
Db      116 EEREKSIEKIWEHLRVVENQCFCGQKKFFG 145

RESULT 11
US-09-542-749A-2
; Sequence 2, Application US/09542749A
; Patent No. 6428981
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Steen T.
; APPLICANT: Christensen, Christina L.
; APPLICANT: Kristensen, Tina
; TITLE OF INVENTION: A Bacillus Protein Production Cell
; FILE REFERENCE: 5861.200-US
; CURRENT APPLICATION NUMBER: US/09/542,749A
; CURRENT FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: 60/130,194
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: PA 1999 00506
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-542-749A-2

Query Match          28.1%; Score 50; DB 4; Length 275;
```

Best Local Similarity 37.0%; Pred. No. 26;
Matches 10; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Qy 7 QLEHRSLSQTLRDQIQRMLFPDEKEFTG 33
Db 91 ELQGRKAGMQFLRNMQESLFVSKKNIT 117

RESULT 12
US-09-252-991A-32259
; Sequence 32259, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32259
; LENGTH: 1201
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32259

Query Match 28.1%; Score 50; DB 4; Length 1201;
Best Local Similarity 39.1%; Pred. No. 1.3e+02;
Matches 9; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Qy 3 LSQEQLEHRSLSQTLRDQIQRML 25
Db 439 VQSQRIQHLEQSLERLQDRRL 461

RESULT 13
US-09-186-276B-56
; Sequence 56, Application US/09186276B
; Patent No. 6388173
; GENERAL INFORMATION:
; APPLICANT: Benfey, Philip
; APPLICANT: Dilaurenzio, Laura
; APPLICANT: Wysocka-Diller, Joanna
; APPLICANT: Malamy, Jocelyn E.
; APPLICANT: Pysh, Leonard
; APPLICANT: Helariutta, Yrjo
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof
; FILE REFERENCE: 5914-075-999
; CURRENT APPLICATION NUMBER: US/09/186,276B
; CURRENT FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 08/842,445
; PRIOR FILING DATE: 1997-04-24
; PRIOR APPLICATION NUMBER: 08/638,617
; PRIOR FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(524)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-186-276B-56

Query Match 27.5%; Score 49; DB 4; Length 524;
Best Local Similarity 30.0%; Pred. No. 73;
Matches 9; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

Qy 5 QEQLHRSLSQTLRDQIQRMLFPDEKEFTG 34
Db 66 QNRVHESENMLNSLRELEKQLLDDDDDESGG 95

RESULT 14
US-08-842-445-56
; Sequence 56, Application US/08842445A
; Patent No. 6441270
; GENERAL INFORMATION:
; APPLICANT: Benfey et al.
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
; FILE REFERENCE: 5914-056-999
; CURRENT APPLICATION NUMBER: US/08/842,445A
; CURRENT FILING DATE: 1997-04-24
; EARLIER APPLICATION NUMBER: 08/638,617
; EARLIER FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Plant
US-08-842-445-56

Query Match 27.5%; Score 49; DB 4; Length 524;
Best Local Similarity 30.0%; Pred. No. 73;
Matches 9; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

Qy 5 QEQLHRSLSQTLRDQIQRMLFPDEKEFTG 34
Db 66 QNRVHESENMLNSLRELEKQLLDDDDDESGG 95

RESULT 15
US-09-186-188B-56
; Sequence 56, Application US/09186188B
; Patent No. 6455672
; GENERAL INFORMATION:
; APPLICANT: Benfey et al.
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses
; FILE REFERENCE: 5914-074-999
; CURRENT APPLICATION NUMBER: US/09/186,188B
; CURRENT FILING DATE: 1998-11-05
; PRIOR APPLICATION NUMBER: 08/842,445
; PRIOR FILING DATE: 1997-04-24
; PRIOR APPLICATION NUMBER: 08/638,617
; PRIOR FILING DATE: 1996-04-26
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 56
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Plant
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(524)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-186-188B-56

Query Match 27.5%; Score 49; DB 4; Length 524;
Best Local Similarity 30.0%; Pred. No. 73;
Matches 9; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

Qy 5 QEQLHRSLSQTLRDQIQRMLFPDEKEFTG 34
Db 66 QNRVHESENMLNSLRELEKQLLDDDDDESGG 95

Search completed: October 15, 2003, 10:32:00

Job time : 9.08171 secs

```

; Sequence 15, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNS
; TITLE OF INVENTION: THERAPEUTIC AN
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/3
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US/09/915
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,500
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 1426
; TYPE: prt
; ORGANISM: Human lgs/bcl9
; US-10-322-579-15

```

```

Query Match      100.0%; Score 178; DB 15; Length 1426;
Best Local Similarity 100.0%; Pred. No. 3.2e-15;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 DGLSQEQLHRRSLQTLRDIQMLFPDEKEFTGA 35
 |||||
 349 DGLSQEQLHRRSLQTLRDIQMLFPDEKEFTGA 383

RESULT 3

```

US-10-322-579-17
; Sequence 17, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMP, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNS
; TITLE OF INVENTION: THERAPEUTIC AN
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/33
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US/09/915
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,50
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 1115
; TYPE: prt
; ORGANISM: Human lgs-1
US-10-322-579-17

```

Query Match 60.1%; Score 107; DB 15; Length 1115;
Best Local Similarity 87.5%; Pred. No. 8.1e-06;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLSQEQLEHRSRSLQTLRDIQRL 25
 |||:|||||:|
 Db 76 GLSKEOLEHRSRSLQTLRDIERLL 99

RESULT 4
US-10-322-579-4
; Sequence 4, Application US/10322579

```

; Sequence 15, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/322,579
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US/09/915,543
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,502
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15
; LENGTH: 1426
; TYPE: PRT
; ORGANISM: Human lgs/bcl9
; US-10-322-579-15

```

```
Query Match      100.0%; Score 178; DB 15; Length 1426;
Best Local Similarity 100.0%; Pred. No. 3.2e-15;
Matches 35; Conservative 0; Mismatches 0; Indels 0;
```

QY 1 DGLSQEQLHRRSLQTLRDIQRLMFLPDEKEFTGA 35
 |||||
 349 DGLSQEQLHRRSLQTLRDIQRLMFLPDEKEFTGA 383

RESULT 3

```

US-10-322-579-17
; Sequence 17, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMP, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNS
; TITLE OF INVENTION: THERAPEUTIC AN
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/33
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US/09/915
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,50
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 1115
; TYPE: prt
; ORGANISM: Human lgs-1
US-10-322-579-17

```

Query Match 60.1%; Score 107; DB 15; Length 1115;
Best Local Similarity 87.5%; Pred. No. 8.1e-06;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLSQEQLEHRSRSLQTLRDIQRL 25
 |||:|||||:|
 Db 76 GLSKEOLEHRSRSLQTLRDIERLL 99

RESULT 4
US-10-322-579-4
; Sequence 4, Application US/10322579

```

; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL
; TITLE OF INVENTION: THERAPEUT
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US
; CURRENT FILING DATE: 2002-12-
; PRIOR APPLICATION NUMBER: US/0
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/2
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Drosophila lgs
; US-10-322-579-4

```

Query Match	37.1%;	Score 66;	DB 15;	Length 35;
Best Local Similarity	31.4%;	Pred. No. 0.05;		
Matches 11;	Conservative 10;	Mismatches 14;	Indels 0;	Gaps 0;

QY 1 DGLSQEQLHRRSLQTLRDIQRMFLFPDEKFTGA 35
 : | : | | | | : : : | | : | |
Dd 1 ENLTPOORQHREEQLAKIKKNQOLFPPENENSVGA 35

RESULT 5

```

US/10-128-714-3511
; Sequence 3511, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,6
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,0
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,6
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,8
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,3
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3511
; LENGTH: 603
; TYPE: PRN
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3511

```

Query Match	32.3%;	Score 57.5;	DB 15;	Length 603;
Best Local Similarity	31.4%;	Pred. No. 18;		
Matches 11;	Conservative 11;	Mismatches 10;	Indels 3;	Gaps 1;

QY 1 DGLSQEQLHRERSLQTLRDIQRMLFPDEKEFTGA 35
||: ||: : : : ||: ||: ||

Db 45 DGVETEKIREKD---EVEKKLERMLFGDDEGFVGA 76

RESULT 6

US-10-128-714-8511
; Sequence 8511, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8511
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8511

Query Match 32.3%; Score 57.5; DB 15; Length 618;
Best Local Similarity 31.4%; Pred. No. 18;
Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1;

QY 1 DGLSQEQLEHRSRSLQTLRDIQRMFLFPDEKEFTGA 35
||: ||: : : : ||||| ||: |||
Db 45 DGVETEKIREKD---EVEKKLERMLFGDDEGFVGA 76

RESULT 7

US-10-053-248-12
; Sequence 12, Application US/10053248
; Publication No. US20030144188A1
; GENERAL INFORMATION:
; APPLICANT: Lin, Biaoyang
; TITLE OF INVENTION: Androgen Regulated Nucleic Acid
; FILE REFERENCE: P-IS 4814
; CURRENT APPLICATION NUMBER: US/10/053,248
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-053-248-12

Query Match 30.9%; Score 55; DB 12; Length 294;
Best Local Similarity 52.2%; Pred. No. 17;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQEQLEHRSRSLQTLRDIQRM 24
|| | | | | : ||| : |||
Db 224 GLRTEGLFRRSASVQTVREIQR 146

RESULT 8

US-10-284-753-32
; Sequence 32, Application US/10284753
; Publication No. US20030129655A1
; GENERAL INFORMATION:
; APPLICANT: Klinger, Tod M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010-1 CIP
; CURRENT APPLICATION NUMBER: US/10/284,753
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 09/507,765
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030129655A1 96572185
US-10-284-753-32

Query Match 30.9%; Score 55; DB 16; Length 333;
Best Local Similarity 52.2%; Pred. No. 19;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQEQLEHRSRSLQTLRDIQRM 24
|| | | | | : ||| : |||
Db 124 GLRTEGLFRRSASVQTVREIQR 146

RESULT 9

US-10-097-340-21
; Sequence 21, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISSEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102

```
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-21
```

```
Query Match      30.9%; Score 55; DB 15; Length 390;
Best Local Similarity 52.2%; Pred. No. 23;
Matches 12; Conservative 4; Mismatches 7; Indels 7; Gaps 0;
```

```
QY      2 GLSQEQLEHRRSLQTLRDIQRM 24
      || || | | | | | | | | | | | |
Db      181 GLRTEGLFRRSASVQTVREIQRL 203
```

```
RESULT 10
US-10-284-753-30
; Sequence 30, Application US/10284753
; Publication No. US20030129655A1
; GENERAL INFORMATION:
; APPLICANT: Klingler, Tod M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010-1 CIP
; CURRENT APPLICATION NUMBER: US/10/284,753
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 09/507,765
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 30
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030129655A1 3068538CD1
US-10-284-753-30
```

```
Query Match      30.9%; Score 55; DB 16; Length 433;
Best Local Similarity 52.2%; Pred. No. 26;
Matches 12; Conservative 4; Mismatches 7; Indels 7; Gaps 0;
```

```
QY      2 GLSQEQLEHRRSLQTLRDIQRM 24
      || || | | | | | | | | | | | |
Db      224 GLRTEGLFRRSASVQTVREIQRL 246
```

```
RESULT 11
US-10-284-753-31
; Sequence 31, Application US/10284753
; Publication No. US20030129655A1
; GENERAL INFORMATION:
; APPLICANT: Klingler, Tod M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010-1 CIP
; CURRENT APPLICATION NUMBER: US/10/284,753
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 09/507,765
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 31
```

```
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030129655A1 404424.5.pseq
US-10-284-753-31
```

```
Query Match      30.9%; Score 55; DB 16; Length 433;
Best Local Similarity 52.2%; Pred. No. 26;
Matches 12; Conservative 4; Mismatches 7; Indels 7; Gaps 0;
```

```
QY      2 GLSQEQLEHRRSLQTLRDIQRM 24
      || || | | | | | | | | | | | |
Db      224 GLRTEGLFRRSASVQTVREIQRL 246
```

```
RESULT 12
US-10-097-340-14
; Sequence 14, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-14
```

```
Query Match      30.9%; Score 55; DB 15; Length 464;
Best Local Similarity 52.2%; Pred. No. 28;
Matches 12; Conservative 4; Mismatches 7; Indels 7; Gaps 0;
```

```
QY      2 GLSQEQLEHRRSLQTLRDIQRM 24
      || || | | | | | | | | | | | |
Db      255 GLRTEGLFRRSASVQTVREIQRL 277
```

RESULT 13

US-10-097-340-16
; Sequence 16, Application US/10097340
; Publication No. US20030087250A1

; GENERAL INFORMATION:

; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT

; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030

; CURRENT APPLICATION NUMBER: US/10/097,340

; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 16

; LENGTH: 643

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-097-340-16

Query Match 30.9%; Score 55; DB 15; Length 643;

Best Local Similarity 52.2%; Pred. No. 41;

Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY

2 GLSQEQLEHRRERSLQTLRDQRM 24

|| || | | | | | | | | | |

434 GLRTEGLFRRRSASVQTVREIQRL 456

Db

RESULT 14

US-10-097-340-19

; Sequence 19, Application US/10097340

; Publication No. US20030087250A1

; GENERAL INFORMATION:

; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY

; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT

; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030

; CURRENT APPLICATION NUMBER: US/10/097,340

; CURRENT FILING DATE: 2002-03-14

; PRIOR APPLICATION NUMBER: 60/276,025

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/325,149

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: 60/276,026

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/324,967

; PRIOR FILING DATE: 2001/09/26

; PRIOR APPLICATION NUMBER: 60/311,732

; PRIOR APPLICATION NUMBER: 60/325,102

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: 60/323,580

; PRIOR FILING DATE: 2001-09-19

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19

; LENGTH: 718

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-097-340-19

Query Match 30.9%; Score 55; DB 15; Length 718;

Best Local Similarity 52.2%; Pred. No. 47;

Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY

2 GLSQEQLEHRRERSLQTLRDQRM 24

|| || | | | | | | | | | |

509 GLRTEGLFRRRSASVQTVREIQRL 531

Db

RESULT 15

US-10-097-340-18

; Sequence 18, Application US/10097340

; Publication No. US20030087250A1

; GENERAL INFORMATION:

; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISEY
; APPLICANT: Peter OLANDT
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. BAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT

; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; FILE REFERENCE: MRI-030

; CURRENT APPLICATION NUMBER: US/10/097,340

; CURRENT FILING DATE: 2002-03-14

; PRIOR APPLICATION NUMBER: 60/276,025

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/325,149

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: 60/276,026


```
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(751)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-097-340-18

Query Match      30.9%; Score 55; DB 15; Length 751;
Best Local Similarity 52.2%; Pred. No. 49;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY      2 GLSQEQLEHERSLQTLRDIQRM 24
      || || | | | | | | | | | |
Db      509 GLRTEGLFRRSASVQTVREIQRL 531

Search completed: October 15, 2003, 10:50:58
Job time : 15.8444 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 8.30739 Seconds
(without alignments)
405.170 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383
Perfect score: 178
Sequence: 1 DGLSQLEHRRSLQTLRDIQRMFLFPDEKEFTGA 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	32.0	1034	2 AB0551	exonuclease SbcC [
2	56	31.5	584	2 S51882	topoisomerase I-re
3	56	31.5	859	2 T46372	hypothetical prote
4	55	30.9	287	2 AE2895	transcription regu
5	55	30.9	295	2 H97670	hex regulon repres
6	55	30.9	643	2 B59436	Rho GTPase activat
7	54.5	30.6	1171	2 S14065	phytochrome B - ri
8	53.5	30.1	4957	2 T03455	ALR protein - huma
9	53.5	30.1	5262	2 T03454	ALR protein - huma
10	53	29.8	227	2 B70438	hypothetical prote
11	53	29.8	376	1 E69957	gamma-D-glutamyl-L
12	53	29.8	818	2 S62790	mismatch DNA recog
13	52.5	29.5	242	2 B70366	hypothetical prote
14	52.5	29.5	705	2 T24343	hypothetical prote
15	52.5	29.5	1039	2 T14802	phytochrome B - so
16	52	29.2	332	2 B47017	probable transcrip
17	52	29.2	332	2 AD2541	transcription init
18	52	29.2	572	2 D82984	pyruvate dehydroge
19	52	29.2	1009	2 S61174	hypothetical prote
20	51	28.7	102	2 AH0216	conserved hypothet
21	51	28.7	237	2 A49940	probable alpha hel
22	51	28.7	237	2 E91056	probable alpha hel
23	51	28.7	237	2 A85901	probable alpha hel
24	51	28.7	329	2 D96834	hypothetical prote
25	51	28.7	477	2 T18801	hypothetical prote
26	51	28.7	899	1 GNMVNM	pol polyprotein -
27	51	28.7	1047	2 G90684	ATP-dependent dsDN
28	51	28.7	1047	2 C85535	ATP-dependent dsDN
29	51	28.7	1464	2 S58984	development protei

30	50.5	28.4	835	2 AD2441	endopeptidase Clp
31	50	28.1	273	2 H69337	conserved hypothet
32	50	28.1	275	2 H69843	hypothetical prote
33	50	28.1	319	2 S49771	hypothetical prote
34	50	28.1	518	2 S86454	CDS protein F9L11.
35	50	28.1	788	2 S67595	hypothetical prote
36	50	28.1	1162	2 D83454	conserved hypothet
37	50	28.1	1236	2 B36329	hypothetical prote
38	49.5	27.8	266	2 AE1124	1-pyrroline-5-carb
39	49.5	27.8	302	1 TPCHTC	troponin T, cardia
40	49.5	27.8	336	2 S72858	hypothetical prote
41	49.5	27.8	830	2 T18860	hypothetical prote
42	49	27.5	310	2 A84142	L-lactate dehydrog
43	49	27.5	338	2 E81730	glycerol-3-phospha
44	49	27.5	420	2 T41236	conserved hypothet
45	49	27.5	537	2 C90796	hypothetical prote

ALIGNMENTS

RESULT 1

AB0551
exonuclease SbcC [imported] - Salmonella enterica subsp. enterica serovar Typhi (str.
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AB0551
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churc
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s.
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AB0551
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1034 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08850.1; PID:g16501663; GSPDB:GN00176
C;Genetics:
A;Gene: STY0429
C;Superfamily: sbcC protein

Query Match 32.0%; Score 57; DB 2; Length 1034;
Best Local Similarity 42.9%; Pred. No. 26;
Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSQLEHRRSLQTLRDIQRMFLFPDEK 30
|: |||: | || | : : | : :
Db 213 LADQLQLEASLQALTDEKRLADQQ 240

RESULT 2

S51882
topoisomerase I-related protein TRP4 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein HRC584; protein O0716; protein YOL115w
C;Species: Saccharomyces cerevisiae
C;Date: 05-May-1995 #sequence_revision 03-Aug-1995 #text_change 21-Jul-2000
C;Accession: S51882; S59158; S58774; S66811
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
submitted to the EMBL Data Library, January 1995
A;Description: Sequence analysis of a 44kb DNA fragment of yeast chromosome XV inclu
and a Delta.
A;Reference number: S51848
A;Accession: S51882
A;Molecule type: DNA
A;Residues: 1-584 <VAN>
A;Cross-references: EMBL:Z48149; NID:g663234; PID:g663237
R;Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
Yeast 11, 1069-1075, 1995
A;Title: Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including
a delta element.

A;Reference number: S59156; MUID:96076631; PMID:7502582
A;Accession: S59158
A;Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A;Residues: 1-584 <VAV>
A;Cross-references: EMBL:Z48149; NID:g663234; PIDN:CAA88145.1; PID:g663237
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
R:Sadoff, B.U.; Heath-Pagliuso, S.; Castano, I.B.; Zhu, Y.; Kieff, F.S.; Christman, M.F.
Genetics 141, 465-479, 1995
A;Title: Isolation of mutants of Saccharomyces cerevisiae requiring DNA topoisomerase I.
A;Reference number: S58774; MUID:96109595; PMID:8647385
A;Accession: S58774
A:Molecule type: DNA
A;Residues: 1-584 <SAD>
A;Cross-references: EMBL:U31355; NID:g950225; PIDN:AAC49091.1; PID:g950226
R:Durand, P.; Hilger, F.; Portetelle, D.; Vandenbol, M.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S66791
A;Accession: S66811
A:Molecule type: DNA
A;Residues: 1-584 <DUR>
A;Cross-references: EMBL:Z74857; NID:g1419986; PID:e251905; PID:g1419987; MIPS:YOL115W
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:TRF4
A;Cross-references: SGD:S0005475; MIPS:YOL115W
A;Map position: 15L
C;Keywords: nucleus

Query Match 31.5%; Score 56; DB 2; Length 584;
Best Local Similarity 36.0%; Pred. No. 19;
Matches 9; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 4 SQEQLEHRRSLQTLRDIQRMFLFPD 28
 |:|::|| | :: |:: |:: |:: |:: |
DB 193 SREEIEIRNQITISTIREAVKQLWPD 217

RESULT 3
T46372
hypothetical protein DKFZp434P1818.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46372
R:Ottewaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23031
A;Accession: T46372
A;Status: preliminary
A:Molecule type: mRNA
A;Residues: 1-859 <AAA>
A;Cross-references: EMBL:AL137528
A;Experimental source: adult testis; clone DKFZp434P1818
C;Genetics:

```

; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kar-
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AE2895
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-287 <KUR>
A;Cross-references: GB:AE008688; PIDN:AA43579.1; PID:gl7741095; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2598
A;Map position: circular chromosome
C;Superfamily: hypothetical protein ybbH

Query Match      30.9%; Score 55; DB 2; Length 287;
Best Local Similarity 40.0%; Pred. No. 11;
Matches 12; Conservative 7; Mismatches 7; Indels 4; Gaps 1;

QY      9 EHRERSLQTLRDIQRMFLF----PDEKEFTG 34
       :|::|::|::|::|::|::|::|::|::|
Db      257 QQRQSMVTLRHRIKQQLVEHRDPDDKQLLG 286

RESULT 5
H97670
hex regulon repressor [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: H97670
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Gold-
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: H97670
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-295 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK88321.1; PID:gl5157797; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4707
A;Map position: circular chromosome
C;Superfamily: hypothetical protein ybbH

Query Match      30.9%; Score 55; DB 2; Length 295;
Best Local Similarity 40.0%; Pred. No. 12;
Matches 12; Conservative 7; Mismatches 7; Indels 4; Gaps 1;

QY      9 EHRERSLQTLRDIQRMFLF----PDEKEFTG 34
       :|::|::|::|::|::|::|::|::|
Db      265 QQRQSMVTLRHRIKQQLVEHRDPDDKQLLG 294

RESULT 6
B59436
Rho GTPase activating protein RhoGAP8 - human
C;Species: Homo sapiens (man)
C;Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 23-Sep-2002
C;Accession: B59436
R;Goward, M.E.; Huckle, E.J.
submitted to GenBank, April 2000
A;Reference number: B59436
A;Accession: B59436
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-643 <GOW>
A;Cross-references: GB:CAB90248; PID:g7711011; PIDN:CAB90248.1

Query Match      30.9%; Score 55; DB 2; Length 643;
Best Local Similarity 52.2%; Pred. No. 28;

```

```
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY      2 GLSQEQLEHRRERSLQTLRD IQRM 24
      || | | | | | | | | | | | |
Db      434 GLRTEGLFRFSASVQTVREIQRL 456

RESULT 7
S14065
phytochrome B - rice
C;Species: Oryza sativa (rice)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C;Accession: S14065
R;Dehesh, K.; Tepperman, J.; Christensen, A.H.; Quail, P.H.
Mol. Gen. Genet. 225, 305-313, 1991
A;Title: phyB is evolutionarily conserved and constitutively expressed in rice seedling
A;Reference number: S14065; MUID:91172131; PMID:2005872
A;Accession: S14065
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1171 <DEH>
A;Cross-references: GB:X57563; NID:g6469490; PIDN:CAA40795.2; PID:g6469491
C;Genetics:
A;Gene: phyB
C;Superfamily: phytochrome; phytochrome homology
C;Keywords: Chromoprotein; photoreceptor; phytochromobilin; transcription regulation
F;103-623/Domain: phytochrome homology <PHYT>
F;364/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match      30.6%; Score 54.5; DB 2; Length 1171;
Best Local Similarity 48.1%; Pred. No. 63;
Matches 13; Conservative 4; Mismatches 5; Indels 5; Gaps 1;

QY      3 LSQEQLEHRRERSLQTLRD IQRM LFPDE 29
      :|| :|| :|| :|| :|| :|| :||
Db      1019 VSQVMQLRERDLQLRDI-----PDE 1040

RESULT 8
T03455
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T03455
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A;Reference number: Z14954; MUID:97388474; PMID:9247308
A;Accession: T03455
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4957 <PRA>
A;Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2358287
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Superfamily: human ALR protein
C;Keywords: alternative splicing

Query Match      30.1%; Score 53.5; DB 2; Length 4957;
Best Local Similarity 42.5%; Pred. No. 4.3e+02;
Matches 17; Conservative 4; Mismatches 12; Indels 7; Gaps 2;

QY      2 GLSQEQLEHRRERSLQTLRD-----IQRM LFPDEKFTGA 35
      |||| :|| :|| :|| :|| :|| :||
Db      2091 GLSQTELE-KQRQRRLRELLIRQQIQRNTLRQKETATAA 2129

RESULT 9
T03454
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
```

```
C;Accession: T03454
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Y
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong hom
A;Reference number: Z14954; MUID:97388474; PMID:9247308
A;Accession: T03454
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5262 <PRA>
A;Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Superfamily: human ALR protein
C;Keywords: alternative splicing

Query Match      30.1%; Score 53.5; DB 2; Length 5262;
Best Local Similarity 42.5%; Pred. No. 4.5e+02;
Matches 17; Conservative 4; Mismatches 12; Indels 7; Gaps 2;
```

```
QY      2 GLSQEQLEHRRERSLQTLRD-----IQRM LFPDEKFTGA 35
      |||| :|| :|| :|| :|| :|| :||
Db      2396 GLSQTELE-KQRQRRLRELLIRQQIQRNTLRQKETATAA 2434
```

```
RESULT 10
B70438
hypothetical protein aq_1596 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: B70438
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70438
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-227 <AQF>
A;Cross-references: GB:AE000747; NID:g2983944; PIDN:AAC07503.1; PID:g2983955; GB:AE0
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_1596
```

```
Query Match      29.8%; Score 53; DB 2; Length 227;
Best Local Similarity 41.4%; Pred. No. 16;
Matches 12; Conservative 7; Mismatches 8; Indels 2; Gaps 1;
```

```
QY      5 QEQLHRRERSLQTLRD IQRM L--FPDEKE 31
      :|| :|| :|| :|| :|| :|| :||
Db      33 KELLEEREKLLLETIRSYEEKLDSFEEEKQ 61
```

```
RESULT 11
E69957
gamma-D-glutamyl-L-diamino acid endopeptid homolog yqgT - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: E69957
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scan
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid
```


Search completed: October 15, 2003, 10:33:10
Job time : 11.3074 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:32 ; Search time 4.22179 Seconds
(without alignments)
389.867 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383
Perfect score: 178
Sequence: 1 DGLSQEHLRSLQTLRDIQRLFPDEKEFTGA 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	178	100.0	1426	1 BCL9_HUMAN	O00512 homo sapien
2	66	37.1	1469	1 BCL9_DROME	Q961d9 drosophila
3	56	31.5	584	1 TRF4_YEAST	P53632 saccharomyc
4	55	30.9	426	1 HEM1_SYNEL	Q8dl53 synechococc
5	55	30.9	718	1 RHG8_HUMAN	Q9nsg0 homo sapien
6	54.5	30.6	174	1 ASH3_MOUSE	Q91jr7 mus musculu
7	54.5	30.6	1171	1 PHYR_ORYSA	P25764 oryza sativ
8	53	29.8	376	1 YOGT_BACSU	P54497 bacillus su
9	53	29.8	817	1 MUTS_THECA	Q9zix6 thermus cal
10	53	29.8	818	1 MUTS_TETH	Q56239 thermus the
11	52.5	29.5	180	1 ASH3_HUMAN	Q9nq33 homo sapien
12	52.5	29.5	190	1 RRF_FUSNN	Q8r5z9 fusobacteri
13	52.5	29.5	242	1 Y755_AQUAE	O66957 aquifex aeo
14	52.5	29.5	1178	1 PHYB_SORBI	P93527 sorghum bic
15	52	29.2	332	1 RPSB_ANASP	Q03065 anabaena sp
16	52	29.2	811	1 MUTS_THEAQ	Q56215 thermus aqu
17	52	29.2	1009	1 RGA2_YEAST	Q06407 saccharomyc
18	51	28.7	237	1 YFHG_ECOLI	P37328 escherichia
19	51	28.7	899	1 POL_MMTVB	P03365 mouse mamma
20	50	28.1	425	1 RHG8_MOUSE	Q9cxp4 mus musculu
21	50	28.1	2004	1 MOZ_HUMAN	Q92794 homo sapien
22	49.5	27.8	301	1 TRT2_CHICK	P02642 gallus gall
23	49.5	27.8	830	1 DYN1_CAEEL	P39055 caenorhabdi
24	49	27.5	310	1 LDH_BACHD	Q9k5z8 bacillus ha
25	49	27.5	328	1 PLSX_STAAN	Q8nx10 staphylococ
26	49	27.5	661	1 YDBJ_SCHPO	Q10369 schizosacch
27	49	27.5	1207	1 DML1_ARATH	Q9sjq6 arabidopsis
28	49	27.5	1234	1 PIP3_HUMAN	Q01970 homo sapien
29	48.5	27.2	184	1 YWKC_BACSU	P45870 bacillus su
30	48.5	27.2	1073	1 CARB_METMA	P58944 methanosarc
31	48.5	27.2	1132	1 PHYB_TOBAC	P29130 nicotiana t
32	48.5	27.2	1505	1 SCP2_RAT	O70608 rattus norv
33	48.5	27.2	1872	1 T2D1_HUMAN	P21675 homo sapien

ALIGNMENTS

RESULT 1

ID	BCL9_HUMAN	STANDARD;	PRT;	1426 AA.
AC	O00512;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	B-cell lymphoma 9 protein (Bcl-9) (legless homolog).			
GN	BCL9.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=98158621; PubMed=9490669;			
RA	Willis T.G.; Zalcberg I.R.; Coignet L.J.A.; Wlodarska I.; Stul M.;			
RA	Jadav D.M.; Bastard C.; Treleaven J.G.; Catovsky D.; Silva M.L.M.;			
RA	Dyer M.J.S.;			
RT	"Molecular cloning of translocation t(1;14)(q21;q32) defines a novel			
RT	gene (BCL9) at chromosome 1q21.";			
RL	Blood 91:1873-1881(1998).			
RN	[2]			
RP	FUNCTION.			
RX	MEDLINE=21952490; PubMed=11955446;			
RA	Kramps T.; Peter O.; Brunner E.; Nellen D.; Froesch B.; Chatterjee S.;			
RA	Murone M.; Zuellig S.; Basler K.;			
RT	"Wnt/wingless signaling requires BCL9/legless-mediated recruitment of			
RT	pygopus to the nuclear beta-catenin-TCF complex.";			
RL	Cell 109:47-60(2002).			
CC	-!- FUNCTION: Involved in signal transduction through the wnt pathway.			
CC	-!- SUBUNIT: Binds to beta-catenin (CTNNB1), PYGO1 and PYGO2.			
CC	-!- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,			
CC	testis, ovary and small intestine, and at lower levels in spleen,			
CC	colon and blood.			
CC	-!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation			
CC	found in a patient with precursor B-cell acute lymphoblastic			
CC	leukemia (ALL). This translocation leaves the coding region			
CC	intact, but may have pathogenic effects due to alterations in the			
CC	expression level of BCL9. Several cases of translocations within			
CC	the 3' untranslated region of BCL9 have been found in B-cell			
CC	malignancies.			
CC	-!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.			
CC	-!- CAUTION: Ref.1 sequence differs from that shown due to a			
CC	frameshift in position 1391.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			

Q99un9 staphylococ
P34934 sus scrofa
P47040 saccharomyc
Q08032 saccharomyc
P43439 enterococcu
O02849 ovis aries
Q8tvb2 methanopyru
P53059 saccharomyc
P44834 haemophilus
Q8ra24 thermoanaer
P58790 agrobacteri
Q9f984 bacillus st

34 48 27.0 328 1 PLSX_STAAM
35 48 27.0 379 1 HS7X_PIG
36 48 27.0 408 1 BTNL_YEAST
37 48 27.0 650 1 CC45_YEAST
38 48 27.0 664 1 NTP1_ENTHR
39 48 27.0 664 1 PDI3_SHEEP
40 47.5 26.7 309 1 PYRB_METKA
41 47.5 26.7 558 1 MNT2_YEAST
42 47.5 26.7 861 1 MUTS_HAEIN
43 47 26.4 184 1 RRF_THETN
44 47 26.4 247 1 HIS4_AGRT5
45 47 26.4 296 1 HSLO_BACST

```
DR EMBL; Y13620; CAA73942.1; ALT_FRAME.
DR Genew; HGNC:1008; BCL9.
DR MIM; 602597; -.
DR GO; GO:0007048; P:oncogenesis; TAS.
KW Nuclear protein; Chromosomal translocation; Proto-oncogene;
KW Wnt signaling pathway.
FT DOMAIN 231 1378 PRO-RICH.
FT DOMAIN 347 377 CTNNB1-BINDING.
FT DOMAIN 331 335 POLY-PRO 1.
FT DOMAIN 514 517 POLY-PRO 2.
FT DOMAIN 900 903 POLY-ALA.
FT DOMAIN 970 973 POLY-PRO 3.
SQ SEQUENCE 1426 AA; 149314 MW; A240A487716B7F1B CRC64;

Query Match 100.0%; Score 178; DB 1; Length 1426;
Best Local Similarity 100.0%; Pred. No. 3.1e-15;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGLSQEQLHRSLSQTLRDIQRLFPDEKEFTGA 35
Db DGLSQEQLHRSLSQTLRDIQRLFPDEKEFTGA 383

RESULT 2
BCL9_DROME STANDARD; PRT; 1469 AA.
ID BCL9_DROME Q9V4D2;
AC Q961D9; Q9V4D2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Bcl-9 homolog (Legless protein).
GN BCL9 OR LGS OR CG2041.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
```

```
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleab J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [3]
RP SEQUENCE OF 6-1469 FROM N.A., AND MUTAGENESIS OF GLY-514; LEU-534 AND
RP ILE-537.
RX MEDLINE=21952490; PubMed=11955446;
RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,
RA Murone M., Zuellig S., Basler K.;
RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
pygopus to the nuclear beta-catenin-TCF complex.";
RL Cell 109:47-60(2002).
CC -!- FUNCTION: Involved in signal transduction through the wnt pathway.
CC -!- SUBUNIT: Binds to ARM and PYGO.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically
throughout development.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF003844; AAF59345.1; ALT_SEQ.
DR EMBL; AY051651; AAK93075.1; -.
DR EMBL; AF457205; AAL91368.1; -.
DR FlyBase; FBgn0039907; lgs.
DR GO; GO:0005634; C:nucleus; IEP.
DR GO; GO:0030528; F:transcription regulator activity; IPI.
DR GO; GO:0030177; P:positive regulation of wnt receptor signali. .; IPI.
DR GO; GO:0007367; P:segment polarity determination; IMP.
KW Nuclear protein; Developmental protein; Segmentation polarity protein;
KW Wnt signaling pathway.
FT DOMAIN 511 555 ARM-BINDING.
FT DOMAIN 1134 1173 ASN-RICH.
FT DOMAIN 1340 1449 GLN-RICH.
FT DOMAIN 1162 1169 POLY-ASN.
FT MUTAGEN 514 514 G->E: IN ALLELE LGS-21L.
FT MUTAGEN 534 534 L->F: IN ALLELE LGS-17E; SEGMENT POLARITY
PHENOTYPE.
FT MUTAGEN 537 537 I->K: IN ALLELE LGS-17P.
SQ SEQUENCE 1469 AA; 153759 MW; 5672E01B7200ED08 CRC64;

Query Match 37.1%; Score 66; DB 1; Length 1469;
Best Local Similarity 31.4%; Pred. No. 1.2;
Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 1 DGLSQEQLHRSLSQTLRDIQRLFPDEKEFTGA 35
Db ENLTPQQRHREQLAKIKMNQFLPENENSVGA 554

RESULT 3
TRF4_YEAST
ID TRF4_YEAST STANDARD; PRT; 584 AA.
AC P53632;
```

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Topoisomerase 1-related protein TRF4.
GN TRF4 OR YOL115W OR O0716 OR HRC584.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96109595; PubMed=8647385;
RA Sadoff B.U., Heath-Pagliuso S., Castano I.B., Zhu Y., Kieff F.S.,
RA Christman M.F.;
RT *Isolation of mutants of Saccharomyces cerevisiae requiring DNA
RT topoisomerase I.;
RL Genetics 141:465-479(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96076631; PubMed=7502582;
RA Vandenbol M., Durand P., Portetelle D., Hilger F.;
RT *Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV
RT including the Tyl-H3 retrotransposon, the sufl(+) frameshift
RT suppressor gene for trna-Gly, the yeast transfer RNA-Thr-1a and a
RT delta element.;
RL Yeast 11:1069-1075(1995).
CC -!- FUNCTION: ESSENTIAL PROTEIN REQUIRED FOR PROPER NUCLEAR DIVISION
CC IN MITOSIS. MAY MEDIATE MITOTIC CHROMOSOME CONDENSATION.
CC -!- SIMILARITY: BELONGS TO THE CID1/TRF4/TRF5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31355; AAC49091.1; -;
DR EMBL; Z48149; CAA88145.1; -;
DR EMBL; Z74857; CAA99134.1; -;
DR PIR; S51882; S51882.
DR SGD; S0005475; TRF4.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IDA.
DR GO; GO:0006265; P:DNA topological change; IGI.
DR GO; GO:0007076; P:mitotic chromosome condensation; IMP.
DR InterPro; IPR002934; NTP_transf.
DR InterPro; IPR001201; PAP_25A_core.
DR InterPro; IPR002058; PAP_assoc.
DR Pfam; PF01909; NTP_transf_2; 1.
DR Pfam; PF03828; PAP_assoc; 1.
KW Mitosis.
SQ SEQUENCE 584 AA; 66030 MW; 8A58B29E4BFDC022 CRC64;

Query Match 31.5%; Score 56; DB 1; Length 584;
Best Local Similarity 36.0%; Pred. No. 8.7;
Matches 9; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 4 SQEQLHRRSLQTLRDIQRLFPD 28
|:|:| |::|:| |::|:|
Db 193 SREEIEIRNQTTISTIREAVKQLWPD 217

RESULT 4
HEM1_SYNEL
ID HEM1_SYNEL STANDARD; PRT; 426 AA.
AC Q8DI53;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GluTR).
GN HEMA OR TLL1738.

OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT *Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.;
RL DNA Res. 9:123-130(2002).
CC -!- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(+) + tRNA(Glu).
CC -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC Involved in chlorophyll biosynthesis.
CC -!- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP005375; BAC09290.1; -;
DR HAMAP; MF_00087; -; 1.
DR InterPro; IPR000343; GluTR.
DR InterPro; IPR000594; ThiF_domain.
DR Pfam; PF00745; GluTR_dimer; 1.
DR Pfam; PF05201; GluTR_N; 1.
DR Pfam; PF05200; GluTR_NAD_bind; 1.
DR TIGRFAMs; TIGR01035; hema; 1.
DR PROSITE; PS00747; GLUTR; FALSE_NEG.
KW Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;
KW NADP; Complete proteome.
FT ACT_SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
FT ACT_SITE 99 99 BASE (BY SIMILARITY).
SQ SEQUENCE 426 AA; 47596 MW; D84CE5A1D2AA777E CRC64;

Query Match 30.9%; Score 55; DB 1; Length 426;
Best Local Similarity 55.6%; Pred. No. 8.2;
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 SQEQLHRRSLQTLRDI 21
|:|:| |::|:| |::|:|
Db 401 SQRDLESQRAMQTLQDL 418

RESULT 5
RHG8_HUMAN
ID RHG8_HUMAN STANDARD; PRT; 718 AA.
AC Q9NSG0; O75983; Q95695; Q96RW1; Q96RW2; Q9HA49; Q9HC46; Q9NVX8;
AC Q9NXL1; Q9UH20;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rho-GTPase-activating protein 8.
GN ARHGAP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT *Novel human cDNA clone with function of inhibiting cancer cell
RT growth.";

RL	Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.	CC	detected are isoforms 1 to 7. Experimental confirmation may be
RN	[2]	CC	lacking for some isoforms;
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).	CC	Name=8;
RA	McDermid H.E., Hu S., Grundy P., Trichet V.;	CC	IsoId=Q9NSG0-1; Sequence=Displayed;
RT	*ARHGAP8: a putative tumor-suppressor gene on chromosome 22q13.3.*;	CC	Name=1;
RL	Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.	CC	IsoId=Q9NSG0-2; Sequence=VSP_001645, VSP_001649, VSP_001652;
RN	[3]	CC	Name=2;
RP	SEQUENCE FROM N.A. (ISOFORMS 5; 6 AND 7).	CC	IsoId=Q9NSG0-3; Sequence=VSP_001645, VSP_001649, VSP_001653,
RC	TISSUE=Colon mucosa, and Mammary gland;	CC	VSP_001655;
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,	CC	Name=3;
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,	CC	IsoId=Q9NSG0-4; Sequence=VSP_001651;
RA	Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,	CC	Name=4;
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,	CC	IsoId=Q9NSG0-5; Sequence=VSP_001650, VSP_001654;
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,	CC	Name=5;
RA	Yamanoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,	CC	IsoId=Q9NSG0-6; Sequence=VSP_001647;
RA	Ninomiya K., Iwayanagi T.;	CC	Name=6;
RT	*NEDO human cDNA sequencing project.*;	CC	IsoId=Q9NSG0-7; Sequence=VSP_001647, VSP_001652, VSP_001656,
RL	Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.	CC	VSP_001657;
RN	[4]	CC	Name=7;
RP	SEQUENCE FROM N.A. (ISOFORM 3).	CC	IsoId=Q9NSG0-8; Sequence=VSP_001646, VSP_001648, VSP_001650,
RA	Goward M.E., Huckle E.J.;	CC	VSP_001654;
RL	Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.	CC	-!- SIMILARITY: Contains 1 CRAL-TRIO domain.
RN	[5]	CC	-!- SIMILARITY: Contains 1 Rho-GAP domain.
RP	SEQUENCE FROM N.A.	CC	-----
RX	MEDLINE=20057165; PubMed=10591208;	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
RA	Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
RA	Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,	CC	the European Bioinformatics Institute. There are no restrictions on its
RA	Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,	CC	use by non-profit institutions as long as its content is in no way
RA	Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,	CC	modified and this statement is not removed. Usage by and for commercial
RA	Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
RA	Clegg S.M., Cobby V.E., Cole C.G., Collier R.E., Connor R.,	CC	or send an email to license@isb-sib.ch).
RA	Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,	CC	-----
RA	Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,	CC	EMBL; AF177331; AAG17975.1; -
RA	Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,	DR	EMBL; AF195968; AAK58136.1; -
RA	Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,	DR	EMBL; AF195969; AAK58137.1; -
RA	Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,	DR	EMBL; AK000192; BAA90999.1; -
RA	Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,	DR	EMBL; AK001306; BAA91614.1; -
RA	Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,	DR	EMBL; AK022305; BAB14008.1; -
RA	Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,	DR	EMBL; AL355192; CAB90248.1; -
RA	McClay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,	DR	EMBL; Z98743; CAB11416.1; ALT_INIT.
RA	Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,	DR	EMBL; Z93244; CAB07531.1; ALT_SEQ.
RA	Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,	DR	EMBL; Z83838; CAB62993.1; ALT_INIT.
RA	Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,	DR	PIR; B59436; B59436.
RA	Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,	DR	HSSP; Q07960; IRGP.
RA	Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,	DR	Genew; HGNC:677; ARHGAP8.
RA	Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,	DR	InterPro; IPR001251; CRAL_TRIO.
RA	Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,	DR	InterPro; IPR000198; RhoGAP.
RA	Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,	DR	Pfam; PF00620; RhoGAP; 1.
RA	Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,	DR	SMART; SM00324; RhoGAP; 1.
RA	Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,	DR	PROSITE; PS50191; CRAL_TRIO; 1.
RA	Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,	DR	PROSITE; PS50238; RHO GAP; 1.
RA	Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,	KW	GTPase activation; Alternative splicing.
RA	Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,	FT	DOMAIN 267 453 CRAL-TRIO.
RA	Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,	FT	DOMAIN 480 666 RHO-GAP.
RA	Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,	FT	VARSPPLIC 1 95 Missing (in isoform 1 and isoform 2).
RA	Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,	FT	/FTid=VSP_001645.
RA	Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,	FT	VARSPPLIC 1 101 Missing (in isoform 7).
RA	Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,	FT	/FTid=VSP_001646.
RA	Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,	FT	VARSPPLIC 1 254 Missing (in isoform 5 and isoform 6).
RA	Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,	FT	/FTid=VSP_001647.
RA	Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,	FT	KIRFYE -> MAPMPT (in isoform 7).
RA	McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelman L.,	FT	/FTid=VSP_001648.
RA	Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,	FT	Missing (in isoform 1 and isoform 2).
RA	Seroussi E., Franssion I., Tapia I., Bruder C.E., O'Brien K.P.,	FT	/FTid=VSP_001649.
RA	Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,	FT	VARSPPLIC 102 107
RA	Tilahun Y., Wright H.;	FT	VARSPPLIC 108 230
RT	*The DNA sequence of human chromosome 22.*;	FT	VARSPPLIC 232 388
RL	Nature 402:489-495(1999).	FT	LORDKAAAALVAVKRPVSVPMAGQDPALSTSHPFYDVA
CC	-!- FUNCTION: GTPase activator for the Rho-type GTPases by converting	FT	RHGILQVAGDDRGRVTVFSCCRMPPSHELDHQLLEYLK
CC	them to an inactive GDP-bound state (By similarity).	FT	YTLQVYVENDYTVIVFYHGLNSRNKPSLGLWLSAYKEFDRK
CC	-!- ALTERNATIVE PRODUCTS:	FT	DGDLTWMVRLVSNKLSLSLPKYWDYRKK -> KRL
CC	Event=Alternative splicing; Named isoforms=8;	FT	LRSSRGDVLAKNPVVRKSYNTPLLNVPQHEAEGAAGG
CC	Comment=Additional isoforms seem to exist. Full isoforms so far	FT	TSIRRHVSSEMTSCPEPQGFSDPPGQGTGTFRRSSPAPHSG
		FT	PCPSRLYPTQPPQGLDPTRRSLRSPENLVQILLESVD
		FT	SDSEGIFIDFGRGRGSGMSDLEGSGGRQSVV (in
		FT	isoform 4 and isoform 7).


```
RESULT 9
MUTS_THECA
ID MUTS_THECA STANDARD; PRT; 817 AA.
AC Q9ZIX6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mutS.
GN MUTS.
OS Thermus caldophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=272;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GK24;
RA Nashiru O., Park B.C., Ko J.H., Kim J.S., Koh S.K., Lee H.C.,
RA Kim C.H., Lee S.Y., Lee D.-S.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a weak ATPase activity (By similarity).
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF007553; AAD01407.1; -.
DR HAMAP; MF_00096; -; 1.
DR InterPro; IPR005748; MutS1.
DR InterPro; IPR000432; MutS_C.
DR InterPro; IPR002863; MutS_N.
DR Pfam; PF01624; MutS_I; 1.
DR Pfam; PF05188; MutS_II; 1.
DR Pfam; PF05192; MutS_III; 1.
DR Pfam; PF05190; MutS_IV; 1.
DR Pfam; PF00488; MutS_V; 1.
DR ProDom; PD001263; MutS_C; 1.
DR SMART; SM00534; MUTSAC; 1.
DR SMART; SM00533; MUTSD; 1.
DR TIGRFAMS; TIGR01070; mutS1; 1.
DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding.
FT NP_BIND 589 596 ATP (POTENTIAL).
SQ SEQUENCE 817 AA; 91179 MW; 61EA066FB84BA761 CRC64;

Query Match 29.8%; Score 53; DB 1; Length 817;
Best Local Similarity 42.3%; Pred. No. 31;
Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 6 EQLEHRERSLQTLRDIQRMFLFPDEKE 31
|:: |::|::| |::| |
Db 464 EKVPQEYRPVQTLKDRQRYTLPEMKE 489

RESULT 10
MUTS_THETH
ID MUTS_THETH STANDARD; PRT; 818 AA.
AC Q56239;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mutS.
GN MUTS.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
```

```
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 1-20, AND CHARACTERIZATION.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=96177563; PubMed=8604304;
RA Takamatsu S., Kato R., Kuramitsu S.;
RT "Mismatch DNA recognition protein from an extremely thermophilic
RT bacterium, Thermus thermophilus HB8.";
RL Nucleic Acids Res. 24:640-647(1996).
RN [2]
RP DOMAIN STRUCTURE.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=98391738; PubMed=9722634;
RA Tachiki H., Kato R., Masui R., Hasegawa K., Itakura H., Fukuyama K.,
RA Kuramitsu S.;
RT "Domain organization and functional analysis of Thermus thermophilus
RT MutS protein.";
RL Nucleic Acids Res. 26:4153-4159(1998).
CC -!- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a maximal ATPase activity at 80 degrees
CC Celsius. Binds double-stranded DNA.
CC -!- SUBUNIT: Homotetramer.
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63810; BAA09880.1; -.
DR HAMAP; MF_00096; -; 1.
DR InterPro; IPR005748; MutS1.
DR InterPro; IPR000432; MutS_C.
DR InterPro; IPR002863; MutS_N.
DR Pfam; PF01624; MutS_I; 1.
DR Pfam; PF05188; MutS_II; 1.
DR Pfam; PF05192; MutS_III; 1.
DR Pfam; PF05190; MutS_IV; 1.
DR Pfam; PF00488; MutS_V; 1.
DR ProDom; PD001263; MutS_C; 1.
DR SMART; SM00534; MUTSAC; 1.
DR SMART; SM00533; MUTSD; 1.
DR TIGRFAMS; TIGR01070; mutS1; 1.
DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding.
FT INIT_MET 0 0
FT DOMAIN 1 129 A1.
FT DOMAIN 130 273 A2.
FT DOMAIN 274 569 B; DNA-BINDING.
FT DOMAIN 570 818 C.
FT NP_BIND 590 597 ATP (POTENTIAL).
SQ SEQUENCE 818 AA; 91249 MW; 657FD676BF033506 CRC64;

Query Match 29.8%; Score 53; DB 1; Length 818;
Best Local Similarity 42.3%; Pred. No. 31;
Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
```

```
QY 6 EQLEHRERSLQTLRDIQRMFLFPDEKE 31
|:: |::|::| |::| |
Db 465 EKVPQEYRPVQTLKDRQRYTLPEMKE 490

RESULT 11
ASH3_HUMAN
ID ASH3_HUMAN STANDARD; PRT; 180 AA.
AC Q9NQ33; Q8WYQ6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Achaete-scute homolog 3 (bHLH transcriptional regulator Sgn-1).
```



```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000706; AAC06923.1; -.
DR PIR; B70366; B70366.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 242 AA; 28618 MW; CE2C25680D721E93 CRC64;

Query Match 29.5%; Score 52.5; DB 1; Length 242;
Best Local Similarity 45.8%; Pred. No. 9.2;
Matches 11; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

QY 10 HRERSLQT-LRDIQRMFLPDEKEF 32
   | : || : |:| ||||
Db 60 HKRTSLRKFFVREIEKMFVFAEKEF 83

RESULT 14
PHYB_SORBI STANDARD; PRT; 1178 AA.
ID PHYB_SORBI STANDARD; PRT; 1178 AA.
AC P93527;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phytochrome B.
GN PHYB OR MA3.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
OX NCBI_TaxID=4558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. 58M;
RX MEDLINE=20188796; PubMed=10723737;
RA Alba R., Kelmenson P.M., Cordonnier-Pratt M.-M., Pratt L.H.;
RT "The phytochrome gene family in tomato and the rapid differential
RL evolution of this family in angiosperms.";
RN Mol. Biol. Evol. 17:362-373(2000).
RP [2]
RC SEQUENCE OF 208-1178 FROM N.A.
RX STRAIN=CV. 58M;
RA MEDLINE=97198556; PubMed=9046599;
RA Childs K.L., Miller F.R., Cordonnier-Pratt M.-M., Pratt L.H.,
RA Morgan P.W., Mullet J.E.;
RT "The Sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a
RL phytochrome B.";
RL Plant Physiol. 113:611-619(1997).
CC -!- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN
CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
CC PROTOCHLOROPHYLLIDE REDUCTASE, RRNA, ETC. IT ALSO CONTROLS THE
CC EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION (BY
CC SIMILARITY).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPHYRROLE CHROMOPHORE.
CC -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
CC -!- SIMILARITY: Contains 1 histidine kinase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF182394; AAB41398.2; -.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR003018; GAF.
DR InterPro; IPR003661; His_kinA.
DR InterPro; IPR005467; His_kinase.
DR InterPro; IPR000014; PAS_domain.
DR InterPro; IPR001294; Phytochrome.
DR Pfam; PF01590; GAF; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR Pfam; PF00512; HSKA; 1.
DR Pfam; PF00989; PAS; 2.
DR Pfam; PF00360; phytochrome; 1.
DR PRINTS; PR01033; PHYTOCHROME.
DR SMART; SM00065; GAF; 1.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00388; HSKA; 1.
DR SMART; SM00091; PAS; 2.
DR TIGRFS; TIGR00229; sensory_box; 2.
DR PROSITE; PS50109; HIS_KIN; 1.
DR PROSITE; PS50112; PAS; 2.
DR PROSITE; PS00245; PHYTOCHROME_1; 1.
DR PROSITE; PS50046; PHYTOCHROME_2; 1.
KW Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW Repeat; Multigene family.
FT DOMAIN 668 739 PAS 1.
FT DOMAIN 802 873 PAS 2.
FT DOMAIN 950 1170 HISTIDINE KINASE.
FT DOMAIN 23 31 POLY-HIS.
FT DOMAIN 43 54 POLY-GLY.
FT BINDING 372 372 CHROMOPHORE (BY SIMILARITY).
SQ SEQUENCE 1178 AA; 129136 MW; C406DF221197B93F CRC64;

Query Match 29.5%; Score 52.5; DB 1; Length 1178;
Best Local Similarity 51.9%; Pred. No. 55;
Matches 14; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 3 LSQEQLEHRRERSLQTLRDIQRMFLPDE 29
   ||| | ||| || ||| |||
Db 1026 VSQAMLLRLRDLQLIRDI-----PDE 1047

RESULT 15
RPSB_ANASP STANDARD; PRT; 332 AA.
AC Q03065;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA polymerase sigma-B factor.
GN SIGB OR ALL7615.
OS Anabaena sp. (strain PCC 7120).
OG Plasmid pCC7120alpha.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054341; PubMed=1385387;
RA Brahamsha B., Haselkorn R.;
RT "Identification of multiple RNA polymerase sigma factor homologs in
RT the cyanobacterium Anabaena sp. strain PCC 7120: cloning, expression,
RT and inactivation of the sigB and sigC genes.";
RL J. Bacteriol. 174:7273-7282(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
```

```
RT cyanobacterium Anabaena sp. strain PCC 7120.*;
RL DNA Res. 8:205-213(2001).
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED.
CC -!- SIMILARITY: Belongs to the sigma-70 factor family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95760; AAA22046.1; -.
DR EMBL; AP003602; BAB77258.1; -.
DR PIR; AD2541; AD2541.
DR HSSP; P00579; 1SIG.
DR InterPro; IPR000943; Sigma_70.
DR Pfam; PF00140; sigma70_r1_2; 1.
DR Pfam; PF04542; sigma70_r2; 1.
DR Pfam; PF04539; sigma70_r3; 1.
DR Pfam; PF04545; sigma70_r4; 1.
DR PRINTS; PR00046; SIGMA70FCT.
DR PROSITE; PS00715; SIGMA70_1; 1.
DR PROSITE; PS00716; SIGMA70_2; 1.
KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
KW DNA-binding; plasmid; Complete proteome.
FT DOMAIN 125 138 POLYMERASE CORE BINDING (POTENTIAL).
FT DNA_BIND 294 313 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 332 AA; 38431 MW; 81323E670CCDF46A CRC64;

Query Match 29.2%; Score 52; DB 1; Length 332;
Best Local Similarity 35.5%; Pred. No. 15;
Matches 11; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 DGLSQEQLEHRSLSQTLRDIQRMFLFPDEKE 31
Db |||||: || |::: ||: ||
248 DGMSPERYAERELLYQDIHNLAKLTPQKE 278
```

Search completed: October 15, 2003, 10:31:02
Job time : 6.22179 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 15, 2003, 10:27:33 ; Search time 20.2918 Seconds
(without alignments)
445.097 Million cell updates/sec

Title: US-09-915-543-15_COPY_349_383
Perfect score: 178
Sequence: 1 DGLSQEQLEHRRSLQTLRDIQRLFPDPKEFTGA 35

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	32.0	1034	16 Q8Z8Y6	Q8z8y6 salmonella
2	57	32.0	1046	16 Q8ZRE3	Q8zre3 salmonella
3	56	31.5	859	4 Q9NT51	Q9nt51 homo sapien
4	56	31.5	1132	4 Q9Y2H2	Q9y2h2 homo sapien
5	56	31.5	1208	5 Q9VXU1	Q9vxu1 drosophila
6	56	31.5	1398	5 Q9VXU2	Q9vxu2 drosophila
7	55	30.9	295	16 Q8UCA0	Q8uca0 agrobacteri
8	55	30.9	426	16 Q8DI53	Q8di53 synechococc
9	55	30.9	433	4 Q8IZM6	Q8izm6 homo sapien
10	55	30.9	818	17 Q8TGZ1	Q8tgz1 methanopyru
11	53.5	30.1	155	11 Q8VD56	Q8vd56 rattus norv
12	53.5	30.1	4957	4 Q14687	Q14687 homo sapien
13	53.5	30.1	5262	4 Q14686	Q14686 homo sapien
14	53	29.8	227	16 Q67532	Q67532 aquifex aeo
15	53	29.8	411	16 Q8PRD9	Q8prd9 xanthomonas
16	52.5	29.5	673	5 Q01302	Q01302 caenorhabdi

17	52	29.2	399	10 Q8W0M3	Q8w0m3 oryza sativ
18	52	29.2	572	16 Q9HTQ7	Q9htq7 pseudomonas
19	52	29.2	616	12 P87708	P87708 fowlpox vir
20	52	29.2	946	4 Q96QE4	Q96qe4 homo sapien
21	52	29.2	1009	3 Q06407	Q06407 saccharomyc
22	52	29.2	1766	12 Q9J599	Q9j599 fowlpox vir
23	52	29.2	2607	11 Q8BTI8	Q8bti8 mus musculu
24	51.5	28.9	1146	10 Q9FPQ2	Q9fpq2 populus tri
25	51.5	28.9	1151	10 Q9FPQ3	Q9fpq3 populus tri
26	51	28.7	102	16 Q8ZFD6	Q8zfd6 yersinia pe
27	51	28.7	108	16 Q8D096	Q8d096 yersinia pe
28	51	28.7	239	16 Q8FF28	Q8ff28 escherichia
29	51	28.7	258	11 Q8BVP0	Q8bvp0 mus musculu
30	51	28.7	329	10 Q8LG66	Q8lg66 arabidopsis
31	51	28.7	329	10 Q9C975	Q9c975 arabidopsis
32	51	28.7	337	5 Q62022	Q62022 caenorhabdi
33	51	28.7	341	4 Q8IYX3	Q8iyy3 homo sapien
34	51	28.7	449	5 Q8IOM4	Q8iom4 caenorhabdi
35	51	28.7	848	16 Q8FMB8	Q8fmb8 corynebacte
36	51	28.7	1047	16 Q8XEJ6	Q8xej6 escherichia
37	51	28.7	1464	5 Q23995	Q23995 drosophila
38	51	28.7	1464	5 Q9VC47	Q9vc47 drosophila
39	51	28.7	1464	5 Q24132	Q24132 drosophila
40	51	28.7	1755	15 Q56220	Q56220 mouse mamma
41	51	28.7	1755	15 Q9IZT3	Q9izt3 exogenous m
42	50.5	28.4	96	5 Q9VN78	Q9vn78 drosophila
43	50.5	28.4	288	13 Q8UVZ1	Q8uvz1 brachydanio
44	50.5	28.4	657	10 Q9FNQ6	Q9fnq6 arabidopsis
45	50.5	28.4	675	4 Q9HD27	Q9hd27 homo sapien

ALIGNMENTS

RESULT 1
Q8Z8Y6 ID Q8Z8Y6 PRELIMINARY; PRT; 1034 AA.
AC Q8Z8Y6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Exonuclease Sbcc.
GN STY0429.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CTL8;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen F.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT *Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CTL8.*;
RL Nature 413:848-852(2001).
DR EMBL; AL627266; CAD08850.1; -.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR004592; Sbcc.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR TIGRFAMS; TIGR00618; sbcc; 1.
KW Complete proteome.
SQ SEQUENCE 1034 AA; 116759 MW; 2513B7573626960A CRC64;

Query Match 32.0%; Score 57; DB 16; Length 1034;
Best Local Similarity 42.9%; Pred. No. 68;

```

Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0
QY 3 LSQEQLEHRERSLQTLRDIQRMFLPDEK 30
   |: |||: | ||| | | : | ||:
Db 213 LADEQLQQLASLQALTDEKRLADQQ 240

RESULT 2
Q8ZRE3 Q8ZRE3 PRELIMINARY; PRT; 1046 AA.
AC Q8ZRE3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ATP-dependent dsDNA exonuclease.
GN SBCC OR STM0395.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
DR EMBL; AE008713; AAL19349.1; -.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR004592; SBCC.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
DR TIGRFAMs; TIGR00618; sbcc; 1.
DR Exonuclease; Complete proteome.
KW SEQUENCE 1046 AA; 117824 MW; BA565CA3BDAD0C82 CRC64;

Query Match 32.0%; Score 57; DB 16; Length 1046;
Best Local Similarity 42.9%; Pred. No. 69;
Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSQEQLEHRERSLQTLRDIQRMFLPDEK 30
   |: |||: | ||| | | : | ||:
Db 213 LADEQLQQLASLQALTDEKRLADQQ 240

RESULT 3
Q9NT51 Q9NT51 PRELIMINARY; PRT; 859 AA.
AC Q9NT51;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP434Pi18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137528; CAB70792.1; -.
DR InterPro; IPR002013; Syja_N.
DR Pfam; PF02383; Syja_N; 1.
DR PROSITE; PS50275; SAC; 1.
KW Hypothetical protein.

```


RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
RA Foslter C., Gabriella A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Hernandez J.R., Houck J.,
RA Harris N.L., Harvey D., Heiman T.J., Howland T.J., Ibegwam C.,
RA Hostin D., Houston K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003499; AAF48467.2; -
DR FlyBase; FBgn0052587; CG32587.
SQ SEQUENCE 1208 AA; 137774 MW; 131F2EB663826A92 CRC64;

Query Match 31.5%; Score 56; DB 5; Length 1208;
Best Local Similarity 45.5%; Pred. No. 1.1e+02;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 3 LSQEQLEHRRERSLQTLRDIQRM 24
Db 811 LQQQQAESQEQASTLRDLRL 832
RESULT 6
Q9VXU2 PRELIMINARY; PRT; 1398 AA.
ID Q9VXU2 Q960D0;
AC Q9VXU2; Q960D0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG32587 protein (SD07366p).
GN CG32587 OR CG6350 OR CG7821.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriella A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003499; AAF48467.2; -
DR FlyBase; FBgn0052587; CG32587.
SQ SEQUENCE 1208 AA; 137774 MW; 131F2EB663826A92 CRC64;

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011625; AAM34916.1; -.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 411 AA; 44337 MW; E193496126355426 CRC64;

Query Match 29.8%; Score 53; DB 16; Length 411;
Best Local Similarity 36.4%; Pred.No. 89;
Matches 12; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 2 GLSQEQLEHRSLSQTLRDIQRMLEFPDEKEFTG 34
| | : || || || : : | | |
DB 17 GASAQSQRETERKQLQRLDELKLTISADRRELEG 49

Search completed: October 15, 2003, 10:35:47
Job time : 23.2918 secs